치의학박사학위논문

Developing a prediction model for soft tissue changes after orthognathic surgeries in skeletal Class III patients based on the partial least squares method

Partial least squares 방법을 이용한
골격성 III급 턱교정수술환자의 연조직 변화 예측

2014년 8월

서울대학교 대학원
치의과학과 치과교정학 전공

이 윤 식

치의학박사학위논문

Developing a prediction model for soft tissue changes after orthognathic surgeries in skeletal Class III patients based on the partial least squares method

지도교수 이 신 재

이 논문을 치의학박사 학위논문으로 제출함

2014년 5월

서울대학교 대학원
치의과학과 치과교정학 전공

이 윤 식

이윤식의 치의학박사 학위논문으로 인준함

2014 년 6 월

위 원 장 (인)

부위원장 (인)

위 원 (인)

위 원 (인)

위 원 (인)

- ABSTRACT -

Developing a prediction model for soft tissue changes after orthognathic surgeries in skeletal Class III patients based on the partial least squares method

Yun-Sik Lee, BS, DDS, MSD

Department of Orthodontics, Graduate School, Seoul National University

*(Directed by Professor **Shin-Jae Lee**, DDS, MSD, PhD, PhD)*

Introduction: The use of bimaxillary surgeries to treat Class III malocclusions makes the results of the surgeries more complicated to accurately estimate. Therefore, our objective is 1) to compare the prediction accuracy of the conventional ordinary least squares (OLS) method with that of the partial least squares (PLS) method; 2) to develop an optimal PLS model for an accurate soft tissue prediction after Class III orthognathic surgery; 3) to compare the prediction performance of the bimaxillary surgery with that of the mandibular surgery.

Material and methods: The subjects of this study consisted of 204 mandibular setback patients who had undergone the combined surgical-orthodontic correction of severe skeletal Class III malocclusions. Among them, 133 patients had maxillary surgeries and 81 patients received additional genioplasties. The prediction model was composed of 226 independent and 64 dependent variables. Two prediction methods, the OLS method and the PLS method were compared. When evaluating the prediction methods, the actual surgical outcome was set as the gold standard. After fitting the equations, test errors were calculated in absolute values and root mean squared values through the leave-one-out cross-validation method.

Results: The validation result demonstrated that the multivariate PLS prediction model with 30 orthogonal components showed the best prediction quality. Using the PLS method, the pattern of prediction errors between 1-jaw surgery and 2-jaw surgery did not show a significant difference. When predicting an anteroposterior soft tissue response after surgery, the vertical components also had a considerable influence on the anteroposterior position and the opposite was also evident.

Conclusions: The multivariate PLS prediction model based on about 30 latent variables might provide an improved algorithm in predicting surgical outcome after 1-jaw or 2-jaw surgical correction for Class III patients

---

Key Words: Class III malocclusion, 2-jaw surgery, profile prediction, partial least squares method

Student Number: 2012-30606

국문초록

Partial least squares 방법을 이용한 골격성 III 급 턱교정수술환자의 연조직 변화 예측

이 윤 식

서울대학교 대학원 치의과학과 치과교정학 전공

(지도교수: 이 신 재)

연구 목적: 본 연구의 목적은 골격성 III 급 부정교합자에 대하여 1) ordinary least squares (OLS) 모형과 partial least squares (PLS) 모형간의 수술 후 연조직 변화 예측의 정확도를 비교하고; 2) 연조직 변화 예측을 위한 최적의 PLS 모형을 개발하며; 3) 양악수술과 하악후퇴수술 사이의 예측 정확도를 비교하는 것이다.

재료 및 방법: 서울대학교치과병원에서 교정 및 악교정 수술을 받은 204 명의 골격성 III 급 부정교합자를 대상으로 하였다. 남자 101 명, 여자 103 명이고 양악수술 대상자는 133 명, 하악후퇴수술 대상자는 71 명이었으며 이부성형술을 시행한 환자는 81 명이다. 양악수술 여부, 부가적 이부성형술 여부, 성별, 비대칭 여부 등의 변인 요소와 술 전 경조직, 연조직 계측점과 술 후 연조직 계측점을 포함하여 226 개의 독립변수가 있으며 64 개의 술후 연조직 좌표(32 개의 계측점)를 종속변수로 하는 예측모형을 수립하였다. 먼저, 통상적인 OLS 모형과 PLS 모형을 만들어 그 예측 정확도를 비교하였다. 정확도는 실제 수술 후의 연조직 모습을 기준으로 하여 각 예측모형의 예측값과 실제값의 차이를 비교하여 구하였다. Training error 는 예측모형을 만든 자료(training dataset)에서의 오차이고, test error 는

예측모형을 만든 자료 이외의 별도의 자료(test dataset)를 예측모형에 적용하여 오차를 구한 것으로 예측모형이 실제 임상에서 얼마나 정확하게 예측을 할 수 있는지 나타낸다. 그리고 실제 환자를 대상으로 하여 OLS 와 PLS 로 예측한 술 후 연조직 모습과 실제 수술 후의 얼굴 사진을 중첩하여 정확도를 직관적으로 비교해 보았다. PLS 모형의 loading graph 를 이용하여 수많은 독립변수 중 어떤 변수들이 술 후 연조직 변화에 주도적으로 영향을 미치는지 살펴보았다. 그리고 RMSEP (root mean squared error of prediction) curve 를 이용하여 최적의 예측모형을 찾는 방법을 살펴보았다. 모형들 간의 정확도를 비교하기 위하여 95% confidence ellipse 를 사용하였다. 마지막으로 앞 단계에서 구한 최적 PLS 모형을 이용하여 양악수술군과 하악후퇴수술이군 사이의 예측 정확도를 비교하였다.

결과: PLS 예측모형이 OLS 예측모형에 비해 통계적으로 유의하게 정확한 예측결과를 보였다. Training error 에서는 OLS 모형과 PLS 모형 모두 정확하게 예측을 하였는데 반하여, test error 에서는 OLS 모형에 비해 PLS 모형이 더 정확한 결과를 보였다. 술 후 연조직 변화를 예측할 때 예측을 하고자 하는 특정한 계측점뿐만이 아니라 주변의 모든 계측점들이 일정한 영향을 주고 있음이 밝혀졌다. 특히 계측점의 수평적인 이동을 예측할 때에 계측점의 수평적 요소뿐만이 아니라 수직적인 요소도 영향을 주는 것으로 나타났다. Training dataset 과 test dataset 에 대하여 각각 RMSEP curve 를 그려서 최적의 PLS 모형을 구하였다. Training dataset 에서는 component 개수가 많을수록 예측오차가 적었지만 test dataset 에서는 component 가 30 개일 때 예측오차가 가장 적었다. 양악수술과 하악후퇴수술의 예측정확도를 비교하였을 때에 코 끝의 수평적 계측점에서는 통계적으로 유의한 차이가 나타났으나 그 이외의 다른 점에서는 양악수술이나 하악후퇴수술 모두 유의한 차이가 없었다.

결론: 1. PLS 방법은 OLS 방법에 비해 예측오차가 작고 더 정확하다.

2. 최적의 PLS 예측모형은 component 가 30 개 일 때의 것이다.

3. 코 끝의 전후방적인 위치를 제외하고는 양악수술과 하악후퇴수술 사이에서 예측정확도 차이를 발견하지 못하였다.

4. PLS 방법이 OLS 방법에 비해 III 급 부정교합자의 술 후 연조직 변화 예측을 더 정확하게 할 수 있다.

주요어: III 급 부정교합, 양악수술, 연조직 변화 예측, partial least squares 방법

학번: 2012-30606

Developing a prediction model for soft tissue changes after orthognathic surgeries in skeletal Class III patients based on the partial least squares method

Yun-Sik Lee, BS, DDS, MSD

Department of Orthodontics, Graduate School, Seoul National University

*(Directed by Professor **Shin-Jae Lee**, DDS, MSD, PhD, PhD)*

Partial least squares 방법을 이용한 골격성 III 급 턱교정수술환자의 연조직 변화 예측

서울대학교 대학원 치의과학과 치과교정학 전공

(지도교수: 이 신 재)

이 윤 식

CONTENTS

Abstract

Contents

Ⅰ. **INTRODUCTION** ……………………………………………………... 1

Ⅱ. **REVIEW OF LITERATURE** ………………………………………. 4

1. Basic mathematics to read this article ……………………………... 4

1.1 Matrix notations ……………………………………………….. 4

1.2 Linear combination and linear model ……………………………. 4

1.3 Orthogonal, linearly independent and uncorrelated variables …………… 5

2. Limitation of the conventional regression model ……………………….. 6

3. Multivariate regression using latent variables ……………………... 7

3.1 Principal component regression or principal component analysis ………... 7

3.2 Partial least squares regression ……………………………………... 9

3.3 Characteristics of PLS in comparison with other methods ……………... 11

Ⅲ. **MATERIALS AND METHODS** ……………………………….. 13

1. Subjects ……………………………………………………………... 13

2. Cephalometric analysis ……………………………………………... 13

3. Variables in predictor and response matrices …………………………….. 14

4. Two multivariate methods to make a prediction equations ………………. 15

5. Inferential statistical analyses ………………………...…………...…..... 15

5.1 Comparison of the prediction accuracy between the OLS method and the PLS method ………………………………………………………….. 15

5.1.1 Training error versus test error ……………………………………….. 15

5.1.2 Goodness-of-fit in the training and test error ………………………… 16

5.1.3 Graphical comparison between the OLS method and the PLS method ... 16

5.2 Development of an optimal PLS model for an accurate soft tissue prediction ……………………………………………………………………….. 17

5.2.1 Analysis of the relationship between predictor variables and soft tissue responses in the PLS method ……………………………………….… 17

5.2.2 Model selection in the PLS method …………………………………... 17

5.3 Comparison of the prediction accuracy between 1-jaw surgery and 2-jaw surgery in Class III patients …………………………………………….. 18

Ⅳ. **RESULTS** ……………………………………………………………... 19

1. Comparison of the prediction accuracy between the OLS method and the PLS method ………………………………………………………………. 19
2. Development of an optimal PLS model for an accurate soft tissue prediction ………………………………………………………………………… 20
3. Comparison of the prediction accuracy between 1-jaw surgery and 2-jaw surgery in Class III patients ………………………………………………. 21

.

Ⅴ. **DISCUSSION** ……………………………………………………………. 23

1. Comparison of the prediction accuracy between the OLS method and the PLS method ………………………………………………………………. 23
2. Development of an optimal PLS model for an accurate soft tissue prediction ………………………………………………………………………… 25
3. Comparison of the prediction accuracy between 1-jaw surgery and 2-jaw surgery in Class III patients ………………………………………………. 27

VI. CONCLUSIONS …………………………………………………….. 29

REFERENCES ……………………………………………………………… 30

APPENDIX TABLE ……………………………………………………… 35

FIGURES ………………………………………………………………….. 36

TABLES …………………………………………………………………... 57

Ⅰ. INTRODUCTION

To make a precise treatment plan and improve patient satisfaction, it is important for orthodontists and surgeons to accurately predict the soft tissue profile changes after orthognathic surgery.[1-3] At present, several software systems provide simulation and prediction of postoperative facial soft tissue changes. Although graphics and user interfaces of the programs have been improved thanks to computer performance enhancement, the prediction results are still far from accurate. The problem is that underlying prediction algorithms based on previous studies are inappropriate to produce sufficient outcomes to use in clinical situations.

A frequently used guide for soft tissue prediction is expressed simply as the 1:1 correspondence ratio for a specific bone to soft tissue change. However, the ratio between the bone and soft tissue changes varies extremely, lacks consistency across studies. For example, the soft-to-hard tissue ratio range was reported to vary from 73%[4] to 100%[3,5] at the lower lip area, 59%[6] to 128%[7] at the soft tissue pogonion.

Correlation analysis and/or regression analysis are also widely used for prediction methods, and they are referred to as the conventional ordinary least squares (OLS) method.[2,3,5,8-11] In particular, it was reported that multiple regression was much more accurate than a simple proportional analysis or a simple regression.[11,12] However, regardless of the number of independent variables incorporated in the multiple regression model, this typical model is unidirectional and univariate, including only a single response (dependent) variable.[13,14] For example, after mandibular setback surgery the soft tissue pogonion will change not only in the anteroposterior dimension but also in vertical dimension, in response to the influence from vertical movement of the adjacent structures. The amounts of surgical (skeletal) movement in the anteroposterior and vertical direction can be included as independent variables at the same time, but the output (response) variable should be only one in the univariate multiple regression model. It is impossible to predict two dimensional variables simultaneously by one equation. A multivariate method is an equation that calculates multiple responses and considers the mutual relationship that may exist among the multiple response variables. In this respect, the multivariate approach, which involves multiple predictors and multiple response variables simultaneously, is more appropriate when predicting a soft tissue response.[13,14]

In addition, OLS method assumes that all predictor variables are independent. But this condition of independence will never exactly be met. For example, a certain degree of vertical skeletal repositioning induces anteroposterior relocation of the soft tissue as well, and vice versa. The soft tissue response at a specific point is not only represented by the underlying bony reference points, but also dependent on its neighboring soft tissue points. Therefore the OLS method has some limitation to predict the precise soft tissue change.[13,14]

The partial least squares (PLS) method is a comparatively new way of formulating prediction equations, and its application to various scientific and biologic disciplines from chemical engineering to brain image analysis is becoming increasingly widespread.[13,15-18] Applying the PLS method is advantageous when the number of variables are many and the variables are highly correlated. The merit of the PLS method is its capability of taking correlation structures into account, not only between the predictor- and response variables, but also controlling for the correlation within the predictor variables and/or the response variables.

Recently, a study applying the multivariate PLS method to mandibular setback surgeries demonstrated considerably more accurate predictions than the OLS method.[12] The conventional OLS method was determined to be unsatisfactory when there were a multitude of correlated variables. Among the variables considered when predicting the soft tissue response to surgery are the patient's age,[2,3,13-16] gender,[2,15-18] time after surgery,[13,19] and pre-surgical soft tissue characteristics including tissue thickness measured at various landmarks.[4,16,20-24] These various factors can be considered in the PLS method through orthogonal linear combinations that are capable of extracting a small number of significant components that are combinations of the original variables.[25] In addition, the improved accuracy of the PLS method is likely due to the fact that the soft tissue response at a specific point is highly dependent on its adjacent soft tissue response, i.e. the interdependency of soft tissue points upon each other.[11]

However, the aforementioned investigation had only been performed for mandibular setback surgeries alone. Two-jaw surgery patients had not yet been included. Previously, the subjects were homogenous in terms of surgical interventions since a homogenous sample was required when investigating the soft tissue response to a specific orthognathic surgery.[14] Including an additional surgery has a great influence on the soft tissue profile change. Most papers have been reporting results of one specific maxillofacial surgical procedure in the sense that the more surgical procedures one adds, the more complex soft tissue prediction becomes.[2,19] There has been an increase in the use of bimaxillary surgery because it is increasingly recognized to produce more stable results than sing-jaw mandibular procedures in Class III correction.[20,21] Prediction programs for bimaxillary surgery were less predictable than for 1-jaw surgery.[6,19,22,23] It would be difficult to exactly determine the changes in the soft tissue profile that are specific to the mandibular setback surgery for other orthognathic surgical procedures, such as Le Fort I osteotomy and/or genioplasty, have been included.

The aim of the present study was 1) to compare the prediction accuracy of the OLS method with that of the PLS method; 2) to develop an optimal PLS model for an accurate soft tissue prediction that can be applied to a various mode of Class III surgical correction: the mandibular surgery and/or the maxillary surgery, and additional genioplasty; 3) to compare the prediction performance of the bimaxillary surgery with that of the mandibular surgery.

Ⅱ. REVIEW OF LITERATURE

1. Basic mathematics to read this article

1.1 Matrix notations

In this paper, the mathematical notations that have been suggested by Geladi and Kowalski[26] are used. Matrices are denoted by upper case bold letter (e.g., $\mathbf{X}$). The identity matrix is denoted by $\mathbf{I}$. All vectors will be column vectors, and denoted by lower case bold letters (e.g., $\mathbf{x}$). Matrix or vector transposition is denoted by an upper case superscript T (e.g., $\mathbf{X}^{\mathrm{T}}$). Two bold letters placed next to each other stand for matrix or vector multiplication. The number of rows, columns, or sub-matrices is denoted by a lowercase italic letter (e.g., i). Predictor variables are stored in an i by j matrix denoted $\mathbf{X}$ whose variable is denoted $\mathbf{x}_{i,j}$ and where the rows are observations and the columns are variables. For convenience, **APPENDIX TABLE** also lists our main notations, acronyms, and terms.

1.2 Linear combination and linear model

A linear combination is an expression constructed from a set of terms by multiplying each term a constant and adding the results. For example, a linear combination of $\mathbf{x}$ and $\mathbf{y}$ would be any expression of the form a$\mathbf{x}$ + b$\mathbf{y}$, where a and b are constants. Definition of linear combination can be written as below.

Suppose that $\boldsymbol{K}$ is a field (i.e., the real number) and $\boldsymbol{V}$ is a vector space over $\boldsymbol{K}$. Then the linear combination of vectors with scalars as coefficients is,

$$a_1\boldsymbol{v_1} + a_2\boldsymbol{v_2} + a_3\boldsymbol{v_3} + \cdots + a_n\boldsymbol{v_n}$$

, where $a_1, \ldots, a_n$ are scalars and $\boldsymbol{v}_1, .., \boldsymbol{v}_n$ are vectors.

The most common example of linear combination in orthodontics is the linear regression model. A general form for the model would be

$$\mathrm{Y} = \beta_0 + \beta_1 \boldsymbol{X_1} + \beta_2 \boldsymbol{X_2} + \beta_3 \boldsymbol{X_3} + \varepsilon$$

, where $\mathbf{X}_i$ are predictors, β_i are unknown parameters, ε is error (or residual) and **Y** is response (*i*=0, 1, 2, 3). It is important to remember that the **parameters enter linearly** in a linear model. The predictors do not have to be linear. For example,

$$\mathbf{Y} = \beta_0 + \beta_1 \boldsymbol{X_1} + \beta_2 \log \boldsymbol{X_2} + \varepsilon$$

is linear but

$$\mathbf{Y} = \beta_0 + \beta_1 \boldsymbol{X_1}^{\beta_2} + \varepsilon$$

is not. Some relationships can be transformed to linearity; $\mathbf{Y} = \beta_0 \boldsymbol{X_1}^{\beta_2}$ can be linearized by taking logs. So, Linear models seem rather restrictive but because the predictors can transformed and combined in any way, they are actually very flexible.[25]

1.3 Orthogonal, linearly independent and uncorrelated variables

Orthogonality is the relation of two lines at right angles to one another. Two vectors, **x** and **y**, are orthogonal if their inner product is zero. Two vector subspaces, **A** and **B**, of an inner product space, **V**, are called orthogonal subspaces if each vector in **A** is orthogonal to each vector in **B**. In other words, orthogonality is the same as perpendicularity.

Orthogonal, linearly independent and uncorrelated are three terms used to indicate lack of relationship between variables. Let **x** and **y** be vector observations of the variables *x* and *y*. Then

1. **x** and **y** are linearly independent iff there exists no constant a such that a**x** -**y** = 0, where **x** and **y** nonnull vectors; iff means if-and-only-if.

2. **x** and **y** are orthogonal iff $\mathbf{x}^{\mathbf{T}}\mathbf{y} = 0$, or $\mathbf{x} \cdot \mathbf{y} = 0$.

3. **x** and **y** are uncorrelated iff $(\mathbf{x} - \bar{x}\mathbf{1})^{\mathrm{T}}(\boldsymbol{y} - \bar{y}\mathbf{1}) = 0$, where $\bar{x}$ *and* $\bar{y}$ are the means of **x** and **y**, respectively, and **1** is vector of ones.

The first important distinction here is that linear independence and orthogonality are properties of the raw variables, while zero correlation is a property of the centered variables. Secondly, orthogonality is a special case of linear independence.

In a geometric view, these terms can be explained as follows.

Each variable is a vector lying in the observation space of n dimensions. Linearly independent variables are those with vectors that do not fall along the same line; that is, there is no multiplicative constant that will expand, contract, or reflect one vector onto the other. Orthogonal variables are a special case of linearly independent variables. Not only do their vectors not fall along the same line, but they also fall perfectly at right angles to one another. The relationship between linear independence and orthogonality is thus straightforward and simple. Meanwhile, to say variables are uncorrelated indicates nothing about the raw variables themselves. Rather, uncorrelated implies that once each variable is centered (i.e., the mean of each vector is subtracted from the elements of that vector), then the vectors are perpendicular. The key to appreciating this distinction is recognizing that centering each variable can and often will change the angle between the two vectors. Thus, orthogonal denotes that the raw variables are perpendicular. Uncorrelated denotes that the centered variables are perpendicular **(Figure 1)**.[27]

2. Limitation of the conventional regression model

The classical regression approach estimates the unknown parameters of the equation using OLS method. The prediction equation can be written as $\mathbf{Y} = \mathbf{X}\mathbf{B}_{\mathrm{OLS}} + \mathbf{E}$. $\mathbf{B}_{\mathrm{OLS}}$ is a $p \times k$ matrix solution of least squares, coefficients $\mathbf{B}_{\mathrm{OLS}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ by multivariate Gauss-Markov theorem **(Figure 2, A)**.[28] Gauss-Markov theorem shows that the least squares estimate $\hat{\mathbf{B}}$ is a good choice when the errors are uncorrelated, normal and have the same variance.

In other words, the OLS method assumes that all the predictor variables are independent, but this condition is not exactly met especially for the numerous dental and facial variables. In multiple regression, collinearities among the predictor variables x_j cause severe problems. The estimated coefficient $\hat{\beta}_j$, can be very unstable and far from their target values. This makes predictions by the regression model to be poor.[29] As regards the prediction of the soft tissue change after orthognathic surgery, skeletal landmarks in a patient are located side by side and move together during the surgical repositioning. Surgery influences all predictor variables. For this reason, the OLS method is not suitable to predict soft tissue change.

Furthermore, the number of predictor variables that can affect soft tissue response, p, is much larger than the number of observations which also makes the effective rank of **X** much smaller than p. When the number of predictors exceeds the number of observations, the likely result will be a model that fits the training dataset perfectly but that will fail to predict test dataset well. This phenomenon is termed *over-fitting*.[18]

3. **Multivariate regression using latent variables**

To solve the problems of multi-collinearity and $n \ll p$ situation, two multivariate regression methods using latent variables were introduced in statistics: principal component regression (PCR) and PLS regression. PCR and PLS can be used with any number of predictor variables, even far more than the number of observations. Both can be better than other methods at forming prediction equations when the standard assumptions of regression analysis are satisfied also.

The latent variable methods is used in order to ‘focus’ the information of large data tables into a few underlying phenomena (also called latent variables, components, or factors) leaving most of the measurement noise behind as residuals. In other words, each object is regarded as a ‘mixture’ of a few underlying phenomena and the aim is to identify and quantify these phenomena with minimal effects of measurement noise.

3.1 Principal component regression or principal component analysis

Principal component analysis (PCA) is one of the most important tools in multivariate statistics. It has been used, for example, in data reduction or visualization of high-dimensional data.[30]

For the two latent variable regression methods, both PCR and PLS, the nonlinear iterative partial least squares (NIPALS) algorithm applies in common. It has been shown that on convergence, the NIPALS solution is the same as that calculated by the eigenvector formulae. The NIPALS algorithm is as follows[26]:

(1) take a vector $\mathbf{x}_j$ from $\mathbf{X}$ and call it $\mathbf{t}_h$: $\mathbf{t}_h = \mathbf{x}_j$
(2) calculate $\mathbf{p}_h^T$: $\mathbf{p}_h^T = \mathbf{t}_h^T\mathbf{X}/\mathbf{t}_h^T\mathbf{t}_h$
(3) normalize $\mathbf{p}_h^T$ to length 1: $\mathbf{p}_{h\ \mathrm{new}}^T = \mathbf{p}_{h\ \mathrm{old}}^T/\|\mathbf{p}_{h\ \mathrm{old}}^T\|$
(4) calculate $\mathbf{t}_h$: $\mathbf{t}_h = \mathbf{X}\mathbf{p}_h/\mathbf{p}_h^T\mathbf{p}_h$
(5) compare the $\mathbf{t}_h$ used in step 2 with that obtained in step 4. If they are the same, stop (the iteration has converged). If they still differ go to step 2.

Then, PCA can be written as a matrix $\mathbf{X}$ of rank r as a sum of r matrices of rank 1: $\mathbf{X} = \mathbf{M}_1 + \mathbf{M}_2 + \mathbf{M}_3 + \ldots + \mathbf{M}_r$. These rank 1 matrices, $\mathbf{M}_h$, can all be written as products of two vectors, a score $\mathbf{t}_h$ and a loading $\mathbf{p}_h^T$: $\mathbf{X} = \mathbf{t}_1\mathbf{p}_1^T + \mathbf{t}_2\mathbf{p}_2^T + \ldots + \mathbf{t}_a\mathbf{p}_a^T$ or the equivalent $\mathbf{X} = \mathbf{T}\mathbf{P}^T$, $\mathbf{P}^T$ is made up of the $\mathbf{p}^T$ as rows and $\mathbf{T}$ of the $\mathbf{t}$ as columns. The latter is a principal component transformation of a data matrix $\mathbf{X}$. This is a representation of $\mathbf{X}$ as its score matrix $\mathbf{T}$. The transformation is $\mathbf{T} = \mathbf{X}\mathbf{P}$ ($= \mathbf{T}\mathbf{P}^T\mathbf{P} = \mathbf{T}\mathbf{I}_n$). So now the OLS style formula can be written as $\mathbf{Y} = \mathbf{T}\mathbf{B}_{\mathrm{PCA}} + \mathbf{E}$ with the solution $\hat{\mathbf{B}} = (\mathbf{T}^T\mathbf{T})^{-1}\mathbf{T}^T\mathbf{Y}$. This regression gives no matrix inversion problems; it is well conditioned.[26] In numerical analysis, this representation of $\mathbf{X}$ as $\mathbf{TB}$ is called singular value decomposition.[31] In short, PCR is a combination of PCA and OLS indeed.

PCR attempts to find linear combinations of the predictors that explain most of the variation in the predictor variables using just a few components. The purpose is dimension reduction. Because the principal components can be linear combinations of all the predictors, the number of variables used is not always reduced. Because the principal components are selected using only the $\mathbf{X}$ matrix and not the response, there is no definite guarantee that the PCR will predict the response particularly well although this often happens. If it happens, we can interpret the principal components in a meaningful way, we may achieve a much simpler explanation of the response. Thus PCR is geared more towards explanation than prediction.

3.2 Partial least squares regression

PLS resembles stepwise multiple linear regression, but in contrast to the latter it is applicable even if the variables are strongly intercorrelated (multi-collinearity situation) and contain significant noise, and even if the number of variables is higher than the number of observations. All variables are included in the final solution; no variables have to be discarded as in stepwise multiple linear regression. That is to say, they allow multiple regressions to be performed without discarding variables, even when both regressors and regressands have noise and the number of objects is low.[32]

The PLS algorithm also resembles canonical correlation between the two matrices, but in contrast to the latter, PLS works even if there are more variables than observations in both matrices, and even if both **X** and **Y** have high noise and multicollinear redundancy.[32]

In case of PCR, $\mathbf{T} = \mathbf{XP}$ ($= \mathbf{TP}^{\mathrm{T}}\mathbf{P} = \mathbf{TI}_n$), $\mathbf{Y} = \mathbf{TB}_{\mathrm{PCR}} + \mathbf{E}$ (solution: $\widehat{\mathbf{B}} = (\mathbf{T}^{\mathrm{T}}\mathbf{T})^{-1}\mathbf{T}^{\mathrm{T}}\mathbf{Y}$) **(Figure 2, B)**. PLS emerged from studies of the flaws in OLS and in PCR. At first, the PLS model was built on the properties of the NIPALS algorithm. The PLS model can be considered as consisting of *outer relations* (**X** and **Y** block individually) and an *inner relation* (linking both blocks). The outer relation for the **X** block is $\mathbf{X} = \mathbf{TP}^{\mathrm{T}} + \mathbf{E} = \mathbf{\Sigma}\ \mathbf{t}_h\mathbf{p}_h^{\mathrm{T}} + \mathbf{E}$. One can build the outer relation for the **Y** block in the same way: $\mathbf{Y} = \mathbf{UQ}^{\mathrm{T}} + \mathbf{F}^* = \mathbf{\Sigma}\ \mathbf{u}_h\mathbf{q}_h^{\mathrm{T}} + \mathbf{F}^*$. It is the intention to describe **Y** as well as is possible and hence to make $||\mathbf{F}^*||$ as small as possible and, at the same time, get a useful relation between **X** and **Y**. The simplest model for this relation is a linear one: $\hat{\mathbf{u}}_h = \mathbf{b}_h\mathbf{t}_h$ where $\mathbf{b}_h = \mathbf{u}_h^{\mathrm{T}}\mathbf{t}_h/\mathbf{t}_h^{\mathrm{T}}\mathbf{t}_h$. An extra loop can be included after convergence to get orthogonal **t** values: $\mathbf{p}^{\mathrm{T}} = \mathbf{t}^{\mathrm{T}}\mathbf{X}/\mathbf{t}^{\mathrm{T}}\mathbf{t}$. The residuals can be calculated from $\mathbf{E}_1 = \mathbf{X} - \mathbf{t}_1\mathbf{p}_1^{\mathrm{T}}$ and $\mathbf{F}_1^* = \mathbf{Y} - \mathbf{u}_1\mathbf{q}_1^{\mathrm{T}}$. In general, $\mathbf{E}_h = \mathbf{E}_{h-1} - \mathbf{t}_h\mathbf{P}_h^{\mathrm{T}}$; $\mathbf{X} = \mathbf{E}_0$. $\mathbf{F}_h^* = \mathbf{F}_{h-1}^* - \mathbf{u}_h\mathbf{q}_h^{\mathrm{T}}$; $\mathbf{Y} = \mathbf{F}_0$. But in the outer relation for the **Y** block, $\mathbf{u}_h$ is replaced by its estimator, $\hat{\mathbf{u}}_h = \mathbf{b}_h\mathbf{t}_h$, and a mixed relation is obtained. The aim is to make $||\mathbf{F}_h||$ small. As a summary, 1) there are outer relations of the form $\mathbf{X} = \mathbf{TP}^{\mathrm{T}} + \mathbf{E}$ and

$\mathbf{Y} = \mathbf{UQ}^T + \mathbf{F}^*$. 2) There is an inner relation $\hat{\mathbf{u}}_h = \mathbf{b}_h \mathbf{t}_h$. 3) The mixed relation is $\mathbf{Y} = \mathbf{TBQ}^T + \mathbf{F}$ where $||\mathbf{F}||$ is to be minimized.[26]

Standard NIPALS PLS algorithm is described below.[31]

A. Centering and normalization

Different normalizations are possible, the specific choice being rather a matter of habit or of convenience. Center and normalize **X** and **Y.**

B. Model fitting

For dimension $h = 1, 2, \ldots, a$.

1) Starting values $u_{ih} = \mathbf{Y}_{ih}$. In case of a single, univariate, **Y** variable ($m = 1$) $\mathbf{u}_h \equiv \mathbf{y}$.

2) Weights for the **X** variables $\mathbf{w}_h = \mathbf{u}_h^T\mathbf{X}$. Normalize **w** to $||\mathbf{w}|| = 1$.

3) Latent variable $\mathbf{t}_h = \mathbf{X}\mathbf{w}_h^T$. In case of a single, univariate, **Y** variable the algorithm continues with step 7, otherwise:

4) Loading for the **Y** variable $\mathbf{b}_y = \mathbf{t}_h^T\mathbf{Y}$.

5) New latent variable values for the **Y** matrix $\mathbf{u}_h = \mathbf{Y}\mathbf{b}_y^T$.

6) If tolerance $||\mathbf{u} - \mathbf{u}_{\text{old}}|| = 10^{-6}||\mathbf{u}||$ then back to step 2, else convergence is reached and step 7 is taken next.

7) Compute the coefficient u_h which relates the latent variable of **X** to the latent variable of **Y**

$$\mathbf{u}_h = u_h \cdot \mathbf{t}_h + d$$

$$u_h = \mathbf{t}_h^T\mathbf{u}_h / ||\mathbf{t}_h||^{1/2}$$

8) Compute the loading for **X**, $\mathbf{b}_{hx}$

$$\mathbf{b}_{hx} = \mathbf{t}_h^T\mathbf{X}$$

9) Compute residuals

$$\mathbf{E}_x = \mathbf{X} - \mathbf{t}_h\, \mathbf{b}_{hx}$$

$$\mathbf{E}_y = \mathbf{Y} - u_h\, \mathbf{t}_h\, \mathbf{b}_{hy}$$

10) For the next dimension, use $\mathbf{E}_x$ instead of **X** and $\mathbf{E}_y$ instead of **Y**. Start again with step 1 above.

C. The number of dimensions, *a*

The number of significant dimensions, *a*, is determined by cross-validation.

D. Prediction error

$\mathbf{e}_i = \mathbf{x}_i$ **-** $\mathbf{x}_i\,\mathbf{B}^T\mathbf{B}$**.** $(\mathbf{x}_i\,\mathbf{B}^T = t_i)$

3.3 Characteristics of PLS in comparison with other methods

There are *hard science* and *soft science* in terms of hardness scale. At the end of the hardness scale are the soft science applications, in which a number, sometimes a very large number, of explanatory variables is available.[33] At first, the PLS method was introduced in the chemical literature as an algorithm[27] and used to be called a soft science. Before late 1980's, it seems that the PLS model was stayed at a heuristic algorithm mainly in chemometrics. PLS has been criticized as an algorithm that solves no well-defined modeling problem.[25] Later, Höskuldsson[34] and Helland[35] interpreted the PLS mathematically, which looked statistically formulated as well.

The PLS method is based on the singular value decomposition of $\mathbf{X}^T\mathbf{Y}$: $\mathbf{X}^T\mathbf{Y} = \sum a_i\mathbf{f}_i\mathbf{g}_i$ where $(\mathbf{f}_i)$ and $(\mathbf{g}_i)$ are orthonormal vectors of appropriate dimension and (a_i) are the singular values arranged in decreasing order. Höskuldsson[34] have shown that upon convergence the weight vectors $\boldsymbol{w}_1$ and $\boldsymbol{q}_1$ correspond to the first pair of left and right singular vectors obtained from a singular vector decomposition (SVD) of the matrix of cross products $\mathbf{X}_0^T\mathbf{Y}_0$. Since the dominant singular value equals $\boldsymbol{w}_1\mathbf{X}_0^T\mathbf{Y}_0\boldsymbol{q}_1 = \boldsymbol{t}_1^T\boldsymbol{u}_1 = (n - 1)\cdot\mathrm{cov}(\boldsymbol{t}_1,\boldsymbol{u}_1)$, the score vectors $\boldsymbol{t}_1$ and $\boldsymbol{u}_1$ have maximum covariance among all score vectors obtainable by applying normalized weights to $\mathbf{X}_0$ and $\mathbf{Y}_0$, respectively.[34] With regard to a more detailed mathematical properties and interpretations, please refer to Höskuldsson[34] and Helland.[35] PLS1 is a univariate version of multivariate PLS, PLS2.[35]

Why can one expect PLS regression methods to perform better than OLS, ridge regression and other well-known regression techniques? The answer is the stability of predictors derived from PLS methods.[34] The essential criterion for the predictability of model is the number of variables included in the models. The uncertainty of the estimated parameters quickly becomes the dominating factor in the variability of predictors. Therefore, it is important to keep the number of variables as low as possible. In PLS components are selected that give 'maximal' reduction in the covariance $\mathbf{X}^T\mathbf{Y}$ of the data. In that sense PLS will give the minimum number of variables that is necessary. Criteria that give penalties on the number of variables, like AIC or BIC, all give rise to more variables than the PLS method.[34]

According to Wold et al.[29] the PLS method is equivalent to the conjugate gradient method used in numerical analysis for related problems. The original algorithm by Wold is essentially another description for the conjugate gradient algorithm for solving the least squares problem with singular **X**. PLS utilizes the principle of dimension reduction by obtaining a small number of latent components that are linear combinations of the original variables to avoid multicollinearity.[36] A special case of response surface is an area where the collinearity problem has been recognized as a serious problem. The procedures of OLS and PCR occupy the opposite ends of a continuous spectrum, with partial least squares lying in between. There are two adjustable 'parameters' controlling the procedure: α, in the continuum [0, 1], and ω, the number of regressors finally accepted. Where α is a real number in the interval [0, 1], with the values 0, ½, and 1 corresponding to OLS, PLS and PCR respectively. The role of α suggests the obvious title for the procedure – 'continuum regression'. These control parameters are chosen by cross- validation.[33]

Compared with other approaches, the PLS algorithm has the followed advantages: the PLS solution is similar to PCR except that the projection **T** is computed *both* to model **X** *and* to correlate with **Y,** while the PCR, **T** is computed only to model **X**. This is accomplished by introducing a weight matrix **W** and a set of latent variables for **Y** denoted by **U** with the corresponding loading matrix $\mathbf{B}_y$. This makes the PLS solution have equal or better predictive properties for **Y**; better in the case when the information in **X** about **Y** appears among the later singular vectors of **X** (corresponding to small singular values).[31] PLS is considered especially useful for constructing prediction equations when there are many explanatory variables and comparatively little sample data.[34]

Ⅲ. MATERIALS AND METHODS

1. Subjects

The subjects consist of 204 patients (103 women, average age 23.8 years; 101 men, average age 23.6 years) who had undergone the surgical correction of a severe Class III malocclusion. All patients were treated at the Department of Orthodontics, and surgery was performed at the Department of Oral and Maxillofacial surgery, Seoul National University Dental Hospital. All patients received mandibular setback surgery and 133 patients had undergone Le Fort I maxillary osteotomy. Additional genioplasty was performed for 81 patients. In all cases, growth had ceased and the patients were healthy and craniofacial deformities or injury was absent.

All patients had been treated with fixed orthodontic appliances before and after surgery. During the preoperative orthodontic treatment, dental decompensation and proper arch coordination was achieved. Postoperative orthodontic treatment was limited to completing the adjustment of the occlusion, and minimal incisor movement was required. The institutional review board for the protection of human subjects reviewed and approved the research protocol (institutional review board no. S-D 20140018).

2. Cephalometric analysis

Lateral cephalograms were taken before and after orthognathic surgery for all patients. Patient was instructed to hold their teeth in occlusion with the lips relaxed when take an X-ray image. Preoperative lateral cephalograms were taken closed to the time of surgical correction. The data were collected prospectively with the postoperative radiographs taken at least 4 months (average 9.1 months) following surgery to allow any residual soft tissue swelling to resolve.[37]

One examiner traced all cephalograms and digitized using a custom program with Microsoft Visual C# (Microsoft, Redmond, USA). To orientate a subject's pre- and postoperative tracings to the same head position, the two tracings were superimposed on the anterior cranial base to confirm whether the sella-nasion planes were coincident. Thirty-nine skeletal landmarks and 32 soft tissue landmarks from glabella to the terminal point were identified. **Figure 3** shows reference planes and cephalometric landmarks used in present study. Capital letters represented the hard tissue landmarks, lowercase letters represented the soft tissue landmarks. With its origin at sella, sella-nasion + 7° was defined as a horizontal reference line, X-axis. sella-nasion is considered to be relatively stable beyond 7 years of age.[20] The vertical reference, Y-axis was established perpendicular to sella-nasion + 7°. The x-y cartesian coordinate system used the liner units in millimeters.

3. Variables in predictor and response matrices

A total of 226 predictor variables (input, explanatory, independent variables, or the **X** matrix) were entered into the prediction equation. The predictor variables included 6 factor variables: patient's age, sex, time after surgery, the amount of facial asymmetry, type of mandibular surgery, existence of maxillary surgery and existence of genioplasty. 78 presurgical skeletal measurements (39 landmarks, measurements in *x*- and *y*-axes), 64 presurgical soft tissue measurements (32 landmarks), and 78 postsurgical skeletal measurements (39 landmarks) also included. The soft tissue changes in the 32 soft tissue landmarks in both the *x* and the *y* axes were included in the 64 response variables (output, dependent variables, or the **Y** matrix).[13,14]

X matrix is a size 204 $(N)_{\text{subjects}}$ × 226 $(K)_{\text{variables}}$ matrix of predictor variables and **Y** matrix is a size 204 $(N)_{\text{subjects}}$ × 64 $(M)_{\text{repsonses}}$ matrix of response variables. As a first step, the mean $\bar{x}_k$ and $\bar{y}_k$ were obtained from the data x_k and y_m, where $_{k = 1, \ldots, K}$ and $_{m = 1, \ldots, M}$ respectively. Then each variable x_k and y_m was scaled to the unit variance. As a result, the $N \times K$ matrix **X** and the $N \times M$ matrix **Y** had the centered and normalized data.[14] The centering makes the following computations numerically well conditioned.[31]

The normalization gives each variable equal influence in the initial stage of the data analysis.

4. Two multivariate methods to make a prediction equations

Two multivariate methods, the conventional OLS method and the PLS method were used to construct prediction equations. OLS is the conventional multivariate linear regression method using forward variable selection coupled with the Akaike information criterion.[38] The prediction equation using the OLS method can be written as $\mathbf{Y} = \mathbf{X}\mathbf{B}_{\mathrm{OLS}} + \mathbf{E}$, where $\mathbf{E}$ is an $N \times M$ matrix of residual for $\mathbf{Y}$, and $\mathbf{B}_{\mathrm{OLS}}$ is a $K \times M$ matrix solution of least squares, coefficient $\mathbf{B}_{\mathrm{OLS}} = (\mathbf{X}^{\mathrm{T}}\mathbf{X})^{-1}\mathbf{X}^{\mathrm{T}}\mathbf{Y}$ using the multivariate Gauss-Markov theorem (**Figure 2, A**).[14,28]

The PLS prediction equation may be written as $\mathbf{Y} = \mathbf{X}\mathbf{B}_{\mathrm{PLS}} + \mathbf{F}$, where $\mathbf{F}$ is an $N \times M$ matrix of residual for $\mathbf{Y}$, and $\mathbf{B}_{\mathrm{PLS}}$ is a $K \times M$ matrix of the PLS prediction coefficients (**Figure 2, B**). In the equation itself, the PLS method resembles the stepwise OLS method. But in contrast to the OLS method, the PLS method is applicable even if the variables are strongly intercorrelated (multicollinear), contain significant noise, and even if the number of variables is greater than the number of subjects (i.e., "small *n* large *p* situations").[32] All predictors are included in the final solution, and no variables have to be discard, which is necessary in the stepwise OLS.[14]

5. Inferential statistical analyses

5.1 Comparison of the prediction accuracy between the OLS method and the PLS method

5.1.1 Training error versus test error

When developing a prediction method, the model is fit for part of the data (the training dataset), and the quality of the fit is judged by how well it predicts the other part of the data (the test dataset, also called the validation dataset). To evaluate the predictive performance of the prediction equations, the leave-one-out cross-validation method was used. Mean training errors and mean test errors were calculated in the OLS method and the PLS method. Mean training errors mean the accuracy of the model on the training dataset which data were used when the model constructed. Mean test errors mean the accuracy of the model on the new data, test dataset which were not used when the model constructed. The mean test errors are much more similar to a real situation **(Figure 4)**.

5.1.2 Goodness-of-fit in the training and test error

After fitting the equation, the *bias* was calculated as a mean difference. The difference between the actual result and predicted position was calculated by subtracting the value for the predicted position from the actual position, $\mathbf{Y_{actual}} - \mathbf{Y_{predicted}}$. Furthermore, the mean absolute error, $|\mathbf{Y_{actual}} - \mathbf{Y_{predicted}}|$ is used as the *criterion of goodness-of-fit* **(Table 2).**

5.1.3 Graphical comparison between the OLS method and the PLS method

Finally, the accuracy of prediction was presented by the actual patient cases. At first, the surgical amounts were calculated by the difference between the lateral cephalograms which were taken as pre-operation and debonding records. Then predictions for soft-tissue change were obtained by V-Ceph (ver 4.3, Osstem, Seoul, Korea), which is a commercial orthodontic analysis and STO (surgical treatment objectives) simulation software and by a custom-made soft-tissue change prediction program. V-Ceph was selected as a tool using the OLS method. Because there is not any software adopting the PLS method, we made our own custom-made prediction software using Microsoft Visual C# (Microsoft, Redmond, USA). Superimposition of lateral facial photo and prediction lines by the OLS method (V-Ceph) and the PLS method were obtained **(Figure 5)**.

5.2 Development of an optimal PLS model for an accurate soft tissue prediction

5.2.1 Analysis of the relationship between predictor variables and soft tissue responses in the PLS method

A PLS prediction model constructed at the previous section was used to analysis the characteristics and relationships the prediction variables. Loading plots were constructed over the predictor variables to identify the effect of the components. Loading plot means the coefficient of the predictor variables. Components are the latent variables that compose the PLS prediction equation. Component 1 is the most powerful component of the prediction model **(Figure 6)**.

Among the 32 soft-tissue landmarks, only four landmarks were selected; pronasale, upper lip, lower lip and pogonion. They are most prominent and clear landmarks that can show the change of soft-tissue of the selected areas. Reconstructed loading plots of which predictor variables were sorted by horizontal and vertical axes were obtained to view the effect of components more easily. Horizontal positions (x-axis) of the landmarks were considered **(Figure 7)**.

5.2.2 Model selection in the PLS method

To fit the model and find the best prediction model, the root mean squared error of prediction (RMSEP) was used as the selection criterion. The RMSEP curve, a graph of RMSEP as a function of the number of components helps us to determine a proper model. In general, it is proposed to use the first local minimum or a deflection points.[39] Model selection was performed on the training dataset and the test dataset respectively **(Figure 8 and 9)**. The aim of the model selection is to find the optimal number of components to minimize the prediction error. Scattergrams and 95% confidence ellipses were constructed to compare the performance of selected model **(Figure 10)**.

5.3 Comparison of the prediction accuracy between 1-jaw surgery and 2-jaw surgery in Class III patients

The PLS prediction model with optimal number of components was chosen. The prediction errors from 1-jaw surgery group and 2-jaw surgery group were compared by *t*-test to identify the difference between the groups **(Table 4)**.

The free statistic software, language R was used.[40] Detailed codes for the multivariate OLS, modified PLS and validation algorithm for use with language R is available by request.

Ⅳ. RESULTS

The subjective characteristics are listed in **Table 1**. The sample size, 204 patients was much larger than previous studies.[13,14] The number of men (101 patients) and the number of women (103 patients) was almost equal. The patients of the subjects had mandibular setback surgery (1-jaw; 71 patients) or bimaxillary surgery (2-jaw; 133 patients) to correct Class III skeletal relationship. A total of 81 patients underwent additional genioplasty. When the patient had maxillary surgery, A point was slightly advanced (1.4 mm) and moved upward (1.1 mm). B point moved backward and upward.

1. Comparison of the prediction accuracy between the OLS method and the PLS method

At the stage of building prediction equations, the goodness of fit or the quality of model fitting can be expressed as the extent of training errors. The training errors from both the OLS and PLS methods were negligible or trivial as shown in **Figure 4, A** and **B**.

The results of the prediction errors after applying the prediction equations in the test dataset from the OLS and PLS methods are summarized in **Table 2-1** (horizontal, *x* position) and **Table 2-2** (vertical, *y* position). After applying the prediction equations in the test dataset, the bias (the error with plus/minus sign) did not show a statistically significant difference between the OLS and PLS methods **(Figure 4, C)** and the mean absolute error showed a statistically significant difference in almost variables **(Figure 4, D)**. Mean absolute error is the average of the absolute value of error, so it means the magnitude of error. PLS method had better prediction performance than OLS method in the view of mean absolute error for almost all variables. However, a comparison test based on the means (i.e., bias) between the predicted and the actual soft tissue profile showed no statistically difference since underestimates and overestimates may cancel each other out.[14,41,42]

For illustrative purpose, four patients were selected to visualize the prediction results between the OLS and PLS methods **(Figure 5)**. Generally, lip and mentolabial fold were the main areas of inaccuracy.[19] The PLS method appeared to perform better than the OLS method in simulating both the mandibular setback and 2-jaw surgeries. Superimposition and comparison to the actual outcome showed that the PLS predictions were closer to the actual outcome than the OLS predictions especially in describing the lip curvatures. As previously reported,[14] the R point and the terminal points were the main areas of inaccuracy in this study.

2. Development of an optimal PLS model for an accurate soft tissue prediction

A loading plot was constructed over the predictor variables to identify the effect of the components **(Figure 6)**. The loading plot visualized the coefficient of the predictor variables. Components were the latent variables that compose the PLS prediction equation and component 1 was the most powerful component of the prediction model. Sex had an peak in the loading plot. This can be explained that among the factor variables, sex of patients had a effect on the soft-tissue prediction model. Repetitive peaks were appeared in the remaining section since the predictor variables, *x* and *y* position of the landmarks were stored by turns.

For convinience, the predictor variables were sorted by anteroposterior (*x*) and vertical (*y*) axes respectively and reconstructed loading plot were obtained. Anteroposterior position of four landmarks; pronasale, upper lip, lower lip and pogonion were chosen (**Figure 7**). As mentioned, sex was the most influential factor variable. It was remarkable that the entire skeletal and soft tissue landmarks had similar influence to the prediction of a specific soft tissue landmark. Although anteroposterior positions of the predictors exercised more influence than the verticla position of the predictors, the vertical positions of the predictors had a certain portion of influence for the soft tissue prediction.

Root mean squared error of prediction (RMSEP) curve was used to select the best prediction model. **Figure 8** showed RMSEP curves in the **training dataset** for selected landmarks. During building a prediction equation in the training dataset, the more components were included, the smaller predictor error were obtained. The full prediction model with the entire components had zero error in the training dataset.

However, when validating the equation to each individual subject, there was an optimum number of components to minimize prediction error in the **test dataset (Figure 9**). **Table 3** showed comparisons of the prediction errors depending on the number of components. The increasing number of the PLS components decreased the prediction errors to some degress, but as the number of components increased more and more, the prediction errors also increased. Typically we choose the smallest model that minimizes the expected prediction error.[17] Therefore, the PLS prediction equation with 30 number of components was selected as the final prediction model.

The prediction performance of a model can be identified by the scattergrams and 95% confidence ellipses.[13,14] The ellipsoid satisfies $(\mathbf{z} - \boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\mathbf{z} - \boldsymbol{\mu}) \leq \chi^2(\alpha)_2$, where $\mathbf{z}$ is the two dimensional (*x*- and *y* coordinates) vector for the error, $\boldsymbol{\mu}$ is the mean vector for $\mathbf{z}$, $\boldsymbol{\Sigma}^{-1}$ is the inverse matrix of the covariance matrix, and $\chi^2(\alpha)_2$ is the upper 95th percentile of a chi-square distribution with two degrees-of-freedom.[28] The contour of an ellipse indicates the 95% confidence boundary. A negative value indicated the prediction was more posterior in the *x*-axis or more superior in the *y*-axis compared to the actual result. If any points are outside the ellipse, they can be called *outliers*.[41] The size of the 95% ellipse for PLS model with optimal number of components (n=30) was smaller than those for minimum and maximum number of components **(Figure 10)**.

3. Comparison of the prediction accuracy between 1-jaw surgery and 2-jaw surgery in Class III patients

Soft tissue prediction accuracy was compared between 1-jaw (mandibular setback) and 2-jawsurgery patients, which is given in **Table 4**. There was no bias in both horizontal and vertical coordination. Horizontal position of supranasal tip and pronasale had a statistically significant difference in mean absolute error and the error from 2-jaw surgery was larger than that of 1-jaw surgery. Except those landmarks, there was no statistically significant difference in the prediction error between 1-jaw surgery and 2-jaw surgery.

Ⅴ. DISCUSSION

In spite of many attempts to predict soft tissue change after orthognathic surgery, there is no appropriate method that is accurate enough to use in clinical situation. There are several difficulties to make these problems. Unexplained individual variations are inevitably present. The responses after surgery across patients are not constant. And statistical and mathematical approaches used to make the prediction model were not suitable to solve the problems.

1. Comparison of the prediction accuracy between the OLS method and the PLS method

In orthodontics, linear regression analysis is the most popular statistic method to develop the prediction equations. A liner regression model was constructed by a **linear combination** of the predictors, a set of terms by multiplying each term a constant and adding the result.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \varepsilon$$

The equation above is the general form of the linear regression model, where X_i are predictors, β_i are unknown parameters, ε is error and Y is response (i=0, 1, 2, 3). It is important to remember that the **parameters enter linearly** in a linear model. The predictors do not have to be linear. For example,

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 \log X_2 + \beta_3 X_3 + \varepsilon$$

is linear but

$$Y = \beta_0 + \beta_1 {X_1}^{\beta_2} + \varepsilon$$

is not. The predictors can be transformed and combined. For example, $Y = \beta_0 {X_1}^{\beta_2}$ can be linearized by taking logs. For this reason, linear regression models are very flexible.

Although a linear model has powerful flexibility to describe reality, there is a fundamental limit. A linear regression analysis is based on the conventional OLS method. The OLS models basically require the prerequisite condition of independence between the predictor variables as well as normality and equality of variance among them. In recent studies, tests for normality and equality of variance were commonly performed to confirm the requisition was satisfied. But the independency between the predictors is seldom considered. Skeletal configuration, dentition and overlaying soft tissue are correlated. Landmarks, angular and linear measurements in cephalometrics also highly correlated. For example when mandibular setback surgery was performed, movement of B point, soft-tissue B point, and nearby soft-tissue landmarks had similar tendency. They could not exist independently. So, correlation between variables in orthodontics is the inherent and fundamental problem. The OLS method cannot solve this problem and that can be the cause of prediction inaccuracy. In this study, there were a number of correlated variables to consider when predicting the soft tissue change following orthognathic surgery; including the patient's age, gender, amount and direction of surgical skeletal movement in the horizontal and vertical direction. The PLS model is the preferred method for constructing a prediction equation when many factors are highly collinear or correlated. Applying the PLS method is even possible when the sample size is less than the number of variables.[39,43]

In this study, we discussed the training error and test error separately. Training error was obtained from the training dataset which was used to construct prediction model. Test error was obtained from the test dataset which was not used to constructed prediction model – a real prediction error for the new data. In other words, the training error can be used as a measure of goodness-of-fit while the test error implies the validity of a prediction model. **Figure 4, A** and **B** showed that the training errors from the two methods were almost null. However, the accuracy of test error in the test dataset is more important than the training error. The OLS method fitted the training dataset perfectly but failed to predict test dataset well **(Figure 4, C** and **D)**. To restate, the OLS method had a disadvantage of *over-fitting*. This phenomenon of the OLS *over-fitting* implies that conventional OLS methods are not satisfactory for complicated soft tissue prediction. On the other hand, the PLS method in this study demonstrated significantly more accurate predictions than the conventional OLS method **(Figure 4, C** and **D**; **Table 2)**.

The PLS prediction showed an improved accuracy when compared to current commercial software programs, Quick Ceph (Quick Ceph Systems, San Diego, USA) and V-Ceph (Osstem, Seoul, Korea) **(Figure 5)**. These two programs provide a one-to-one soft tissue ratio setting for the movement of the corresponding hard tissue. Although the exact algorithms for these programs are unknown and would be confidential, aforementioned settings reflect the application of the simple OLS method as their algorithm.

Most of the reported inaccuracies in the soft tissue predictions were the upper lip,[5,44] lower lip,[22,23,45-48] and mentolabial fold.[6,19,49] This was not the case for the PLS results, as shown. Although the results of the PLS predictions in this study were not perfect, the improved accuracy seems obvious **(Figure 4)**.

Linearly independent, orthogonal and uncorrelated are three terms used to indicate lack of relationship between variables. Linearly independent variables are those with vectors that do not fall along the same line; there is no multiplicative constant that will expand, contract, or reflect one vector onto the other. Orthogonal variables are a special case of linearly independent variables. Not only do their vectors not fall along the same line, but they also are at right angles to one another. Uncorrelated implies that once each variable is centered (i.e., the mean of each vector is subtracted from the elements of that vector), then the vectors are perpendicular. Orthogonal denotes that the raw variables are perpendicular. Uncorrelated denotes that the centered variables are perpendicular.[50] It is important to remember these concepts because the conventional OLS method which is the most familiar statistic method to us assumes that variables of the prediction model are independent each other.

2. **Development of an optimal PLS model for an accurate soft tissue prediction**

We explored the complex relationship between predictor variables and soft tissue responses by invoking an intricate multivariate statistical analysis. As an example, to identify factors that might influence the soft tissue response, we depicted loading plots for the anteroposterior landmarks; pronasale, upper lip, lower lip and pogonion. As depicted in **Figure 7**, the loading values indicate the magnitude of each predictor variable in predicting the response. The loading values are useful not only in determining the influence of each variable but also to develop computer algorithms. The loading pattern showed that several predictor variables had higher values of influence than others. For example, the sex predictor variable had an important role among factor variables. Consistent with the previous report, this finding signifies that soft tissue movement in response to skeletal repositioning is somewhat greater in females than in males.[5]

Three additional conclusions may also be plausibly drawn from the loading plot. 1) Both the pre-surgical skeletal and soft tissue characteristics as well as the amount of surgical repositioning contributed to predicting the soft tissue response after surgery. 2) When predicting a soft tissue response in the x-axis, anteroposterior variables had higher loading values than vertical predictor variables. However, the vertical predictor variables did not have a minor or trivial role but have considerable influence on the anteroposterior outcome to some extent. 3) It was also notable that some neighboring soft tissue landmarks and all the skeletal landmarks as a whole had a greater influence on the predictions of specific soft tissue landmarks then the pre-surgical landmark of that actual soft tissue point itself (arrow of a specific pre-treatment landmark in **Figure 7**). The complexity of these relationships is the reason overly simplistic conventional OLS predictions or simple 1-to-1 ratio statistics upon which current software programs have depended demonstrate lesser accuracy.

After finding the superiority of the PLS method over the OLS method, the next step was choosing the best prediction model. The best prediction model can be defined as the simplest model that minimizes the test error. For the model selection criteria, a squared type of error, namely the root mean squared error of prediction (RMSEP), was used. The RMSEP has been frequently used to assess the prediction performance and to choose the optimal number of components in principal components regression.[51,52] RMSEP can be obtained from cross-validation method or ordinary bootstrap estimate. Mevik and Cederkvist reported that leave-one-out cross-validation was preferable when there were more variables than observations.[52] In this study, RMSEP curves obtained from leave-one-out cross-validated prediction were used.

Figure 8 and **Figure 9** showed the difference in the trend of optimal number of components between the training dataset and the test dataset. In the training dataset, the more components were included, the smaller predictor error were obtained. But in the test dataset, too many components induced lager prediction error. More complex prediction model which have an excessive number of parameters fits better in the training dataset but not in the test dataset as it may exaggerate minor flucutations in the data. It is *over-fitting*, as mentioned above. In order to avoid over-fitting, cross-validation is useful. In this study, we used leave-one-out cross-validation.

The 95% confidence ellipse is a good menas to compare the accuarcy of the 2-D landmark. This shows the confidence interval of *x*- and *y*- coordinate at a time. **Figure 10** showed that the PLS prediction model with the optimal number of components had smaller prediction error than the prediction moel with minimum and maximum number of components. The magnitude of the prediction error of *x*- and *y*- coordinate was different. This can be explained as the definition of the landmark. For example, pronasale is the most anterior point of the nasal tip, so it is easier to the anteroposterior position exactly than the vertical position of the pronasale. It is noticeable that the shape of the confidence ellipse of the optimal number of components are almost circular. The optimal prediction model may overcome the weak point of prediction accuarcy by the landmark definition.

3. **Comparison of the prediction accuracy between 1-jaw surgery and 2-jaw surgery in Class III patients**

Previous studies were not able to analyze and interpret more than one type of surgery or more than one vector of movement once at a time of investigation. Since conventional methods could not properly handle the complex data structures, only identical surgical procedures could be analyzed. Isolated mandibular prognathism occurs in a relatively small portion of Class III patients.[3,21] Therefore, the combination of a Le Fort I osteotomy of the maxilla and a mandibular setback surgery seems to be the current trend for skeletal Class III treatment.[20,53] With previous prediction methods, even an additional genioplasty was considered as a confounding variable when predicting soft tissue responses.[2] For those patients who undergo maxillary surgery and/or genioplasty, the vectors of movement are not uniform. Thus patients undergoing and additional jaw surgery would have less predictable results than those undergoing a relatively simple mandibular setback surgery.[19,22,23] Even though, the prediction of 2-jaw surgery has a potential to get larger error, in this study's use of the PLS method there was no statistically significant difference in the prediction error between 1-jaw surgery and 2-jaw surgery, either in vertical or anteroposterior direction except anteroposterior position of supranasal tip and pronasale. The type of surgery may influenced the anteroposterior change of the nose (supranasal tip and pronasale). But except those points, the type of surgery did not influenced on the accuracy of prediction **(Table 4)**.

The multivariate PLS method has been improved rapidly with the advent of high-speed computers. Computer assisted predictions have become an integral part of surgical-orthodontic treatment planning.[46] However, existing software predictions still result in considerable errors. One of the reasons for the errors might be caused by over-simplistic OLS algorithms that were integrated into most commercially available computer programs. These programs have never been clearly published or opened to the public. We hope that the soft tissue prediction method presented in this study will provide a practical algorithm to improve surgical treatment simulation programs.

VI. CONCLUSIONS

1. It was our observation that the more sophisticated PLS mathematical method predicted better than the conventional OLS method, and by applying the multivariate PLS method, prediction errors can be minimized or possibly completely eliminated.

2. Among the methods and variables tested, the multivariate PLS prediction model based on about 30 latent variable components showed the best prediction quality.

3. There is no statistically significant difference in the prediction error between 1-jaw surgery group and 2-jaw surgery group except anteroposterior change of nasal tip area (supranasal tip and pronasale).

4. Based on our findings, we propose that the PLS method might provide an improved algorithm in predicting surgical outcomes after mandibular setback or bimaxillary surgical correction for Class III patients.

REFERENCES

1. Joss CU, Joss-Vassalli IM, Kiliaridis S, Kuijpers-Jagtman AM. Soft tissue profile changes after bilateral sagittal split osteotomy for mandibular advancement: a systematic review. J Oral Maxillofac Surg 2010;68:1260-1269.
2. Joss CU, Joss-Vassalli IM, Berge SJ, Kuijpers-Jagtman AM. Soft tissue profile changes after bilateral sagittal split osteotomy for mandibular setback: a systematic review. J Oral Maxillofac Surg 2010;68:2792-2801.
3. Joss CU, Vassalli IM, Thuer UW. Stability of soft tissue profile after mandibular setback in sagittal split osteotomies: a longitudinal and long-term follow-up study. J Oral Maxillofac Surg 2008;66:1610-1616.
4. Alves PV, Mazucheli J, Vogel CJ, Bolognese AM. How the lower face soft tissue changes after mandibular advancement or setback. J Craniofac Surg 2008;19:593-598.
5. Mobarak KA, Krogstad O, Espeland L, Lyberg T. Factors influencing the predictability of soft tissue profile changes following mandibular setback surgery. Angle Orthod 2001;71:216-227.
6. Enacar A, Taner T, Toroglu S. Analysis of soft tissue profile changes associated with mandibular setback and double-jaw surgeries. Int J Adult Orthodon Orthognath Surg 1999;14:27-35.
7. Schatz JP, Tsimas P. Cephalometric evaluation of surgical-orthodontic treatment of skeletal Class III malocclusion. Int J Adult Orthodon Orthognath Surg 1995;10:173-180.
8. Ingervall B, Thuer U, Vuillemin T. Stability and effect on the soft tissue profile of mandibular setback with sagittal split osteotomy and rigid internal fixation. Int J Adult Orthodon Orthognath Surg 1995;10:15-25.
9. Chew MT, Sandham A, Wong HB. Evaluation of the linearity of soft- to hard-tissue movement after orthognathic surgery. Am J Orthod Dentofacial Orthop 2008;134:665-670.
10. Ksiezycki-Ostoya BK, McCollum AGH, Becker PJ. Sagittal soft-tissue changes of the lower lip and chin associated with surgical maxillary impaction and consequent

mandibular autorotation. Semin Orthod 2009;15:185-195.

11. McCollum AGH, Gardener GJM, Evans WG, Becker PJ. Soft-tissue changes related tomandibular advancement surgery. Semin Orthod 2009;15:161-171.
12. Kneafsey LC, Cunningham SJ, Petrie A, Hutton TJ. Prediction of soft-tissue changes after mandibular advancement surgery with an equation developed with multivariable regression. Am J Orthod Dentofacial Orthop 2008;134:657-664.
13. Lee HJ, Suh HY, Lee YS, Lee SJ, Donatelli RE, Dolce C, et al. A better statistical method of predicting postsurgery soft tissue response in Class II patients. Angle Orthod 2014;84:322-328.
14. Suh HY, Lee SJ, Lee YS, Donatelli RE, Wheeler TT, Kim SH, et al. A more accurate method of predicting soft tissue changes after mandibular setback surgery. J Oral Maxillofac Surg 2012;70:e553-562.
15. Krishnan A, Williams LJ, McIntosh AR, Abdi H. Partial Least Squares (PLS) methods for neuroimaging: a tutorial and review. Neuroimage 2011;56:455-475.
16. Lee SJ. Modified partial least squares method incorporating mixed effect model. Department of Statistics, Graduate School. Seoul, Korea: Korea University; 2012: p. 1-50.
17. Wehrens R. Chemometrics with R: Multivariate Data Analysis in the Natural Sciences and Life Sciences. Heidelberg: Springer; 2011.
18. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. Data mining, inference, and prediction. New York: Springer Verlag; 2009.
19. Jones RM, Khambay BS, McHugh S, Ayoub AF. The validity of a computer-assisted simulation system for orthognathic surgery (CASSOS) for planning the surgical correction of class III skeletal deformities: single-jaw versus bimaxillary surgery. Int J Oral Maxillofac Surg 2007;36:900-908.
20. Johnston C, Burden D, Kennedy D, Harradine N, Stevenson M. Class III surgical-orthodontic treatment: a cephalometric study. Am J Orthod Dentofacial Orthop 2006;130:300-309.
21. Bailey LJ, Cevidanes LH, Proffit WR. Stability and predictability of orthognathic surgery. Am J Orthod Dentofacial Orthop 2004;126:273-277.

22. Eckhardt CE, Cunningham SJ. How predictable is orthognathic surgery? Eur J Orthod 2004;26:303-309.
23. Kaipatur NR, Flores-Mir C. Accuracy of computer programs in predicting orthognathic surgery soft tissue response. J Oral Maxillofac Surg 2009;67:751-759.
24. Geladi P, Kowalski BR. An example of 2-block predictive partial least-squares regression with simulated data. Anal Chim Acta 1986;185:19-32.
25. Faraway JJ. Practical regression and ANOVA using R. 2002.
26. Geladi P, Kowalski BR. Partial least-squares regression - a tutorial. Anal Chim Acta 1986;185:1-17.
27. Bisani ML, Faraone D, Clementi S, Esbensen KH, Wold S. Principal components and partial least-squares analysis of the geochemistry of volcanic-rocks from the Aeolian Archipelago. Anal Chim Acta 1983;150:129-143.
28. Johnson RA, Wichern DW. Applied multivariate statistical analysis. Pearson Prentice Hall; 2007.
29. Wold S, Ruhe A, Wold H, Dunn WJ. The collinearity problem in linear-regression - the partial least-squares (pls) approach to generalized inverses. Siam J Sci Stat Comp 1984;5:735-743.
30. Lee D, Lee W, Lee Y, Pawitan Y. Super-sparse principal component analyses for high-throughput genomic data. Bmc Bioinformatics 2010;11:296.
31. Lindberg W, Persson JA, Wold S. Partial least-squares method for spectrofluorimetric analysis of mixtures of humic-acid and ligninsulfonate. Anal Chem 1983;55:643-648.
32. Martens M, Martens H, Wold S. Preference of cauliflower related to sensory descriptive variables by partial least-squares (pls) regression. J Sci Food Agr 1983;34:715-724.
33. Stone M, Brooks RJ. Continuum regression - cross-validated sequentially constructed prediction embracing ordinary least-squares, partial least-squares and principal components regression. J Roy Stat Soc B Met 1990;52:237-269.
34. Höskuldsson A. PLS regression methods. J Chemometr 1988;2:211-228.
35. Helland IS. Partial least-squares regression and statistical-models. Scand J Stat 1990;17:97-114.

36. Chun H, Keles S. Sparse partial least squares regression for simultaneous dimension reduction and variable selection. J Roy Stat Soc B 2010;72:3-25.
37. Dolce C, Van Sickels JE, Bays RA, Rugh JD. Skeletal stability after mandibular advancement with rigid versus wire fixation. J Oral Maxillofac Surg 2000;58:1219-1227; discussion 1227-1218.
38. Akaike H. A new look at the statistical model identification. IEEE Trans Autom Contr 1974;19:716-723.
39. Mevik BH, Wehrens R. The pls package: principal component and partial least squares regression in R. J Stat Softw 2007;18:1-24.
40. R Development Core Team. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing; 2014.
41. Donatelli RE, Lee SJ. How to report reliability in orthodontic research: Part 1. Am J Orthod Dentofacial Orthop 2013;144:156-161.
42. Donatelli RE, Lee SJ. How to report reliability in orthodontic research: Part 2. Am J Orthod Dentofacial Orthop 2013;144:315-318.
43. Tobias RD. An introduction to partial least squares. TS-509. Cary, NC: SAS Institute Inc.; 2006.
44. Kasai K. Soft tissue adaptability to hard tissues in facial profiles. Am J Orthod Dentofacial Orthop 1998;113:674-684.
45. Naoumova J, Soderfeldt B, Lindman R. Soft tissue profile changes after vertical ramus osteotomy. Eur J Orthod 2008;30:359-365.
46. Pektas ZO, Kircelli BH, Cilasun U, Uckan S. The accuracy of computer-assisted surgical planning in soft tissue prediction following orthognathic surgery. Int J Med Robot 2007;3:64-71.
47. Sameshima GT, Kawakami RK, Kaminishi RM, Sinclair PM. Predicting soft tissue changes in maxillary impaction surgery: a comparison of two video imaging systems. Angle Orthod 1997;67:347-354.
48. Donatsky O, Bjorn-Jorgensen J, Hermund NU, Nielsen H, Holmqvist-Larsen M, Nerder PH. Immediate postoperative outcome of orthognathic surgical planning, and prediction of positional changes in hard and soft tissue, independently of the extent

and direction of the surgical corrections required. Br J Oral Maxillofac Surg 2011;49:386-391.

49. Power G, Breckon J, Sherriff M, McDonald F. Dolphin Imaging Software: an analysis of the accuracy of cephalometric digitization and orthognathic prediction. Int J Oral Maxillofac Surg 2005;34:619-626.

50. Rodgers JL, Nicewander WA, Toothaker L. Linearly independent, orthogonal, and uncorrelated variables. Am Stat 1984;38:133-134.

51. Martens H, Næs T. Multivariate Calibration. Chichester: Wiley; 1989.

52. Mevik BH, Cederkvist HR. Mean squared error of prediction (MSEP) estimates for principal component regression (PCR) and partial least squares regression (PLSR). J Chemom 2004;18:422-429.

53. Jung MH. Age, extraction rate and jaw surgery rate in Korean orthodontic clinics and small dental hospitals. Korean J Orthod 2012;42:80-86.

APPENDIX TABLE. List of notations, acronyms, terms used in this study

Notation	
N	the number of samples or observations
p	the number of predictor (x) variables
k	the number of response (y) variables
a	the number of latent factors used (≤ rank of **X**)
r	the number of levels in the random variable
$\boldsymbol{x}$	a column vector of features for the predictor variables (size $p \times 1$)
$\boldsymbol{y}$	a column vector of features for the response variables (size $k \times 1$)
$\mathbf{X}$	a matrix of features for the predictor variables (size $n \times p$)
$\mathbf{Y}$	a matrix of features for the response variables (size $n \times k$)
$\mathbf{b}$	a column vector of coefficients
$\mathbf{B}$	a matrix of coefficients for the multivariate methods (size $p \times k$)
$\mathbf{t}_h$	a column vector of scores for the **X** block, factor h (size $n \times 1$)
$\mathbf{p}_h^{\mathrm{T}}$	a row vector of loadings for the **X** block, factor h (size $1 \times p$)
$\mathbf{w}_h^{\mathrm{T}}$	a row vector of weights for the **X** block, factor h (size $1 \times p$)
$\mathbf{T}$	the matrix of **X** scores (size $n \times a$)
$\mathbf{P}^{\mathrm{T}}$	the matrix of **X** loadings (size $a \times p$)
$\mathbf{u}_h$	a column vector of scores for the **Y** block, factor h (size $n \times 1$)
$\mathbf{q}_h^{\mathrm{T}}$	a row vector of loadings for the **Y** block , factor h (size $1 \times k$)
$\mathbf{U}$	the matrix of scores (size $n \times a$)
$\mathbf{Q}^{\mathrm{T}}$	the matrix of **Y** loadings (size $a \times k$)
$\mathbf{M}_h$	a rank 1 matrix, outer product of $\mathbf{t}_h$ and $\mathbf{p}_h^{\mathbf{T}}$ (size $n \times p$)
$\mathbf{E}_h$	the residual of **X** after subtraction of h components (size $n \times p$)
$\mathbf{F}_h$	the residual of **Y** after subtraction of h components (size $n \times k$)
$\mathbf{b}_h$	the regression coefficient for one PLS component
$\mathbf{I}_n$	the identity matrix of size $n \times n$
Acronyms	
OLS	ordinary least squares regression, multiple multivariate linear regression
PCR	principal components regression
PLS	partial least squares regression
PLSM	modified PLS method implementing mixed effect model
CV	cross-validation
Terms	
Predictor variable	also called input-, explanatory-, descriptive-, regressor-, or independent variables
Response variable	also called output-, regressand-, or dependent variables
Training dataset	also called learning-, calibration-, or study dataset
Test dataset	also called real-, prediction-, or validation dataset
Component	also called latent variable or latent factor
Coefficient	also called loading or sensitivity

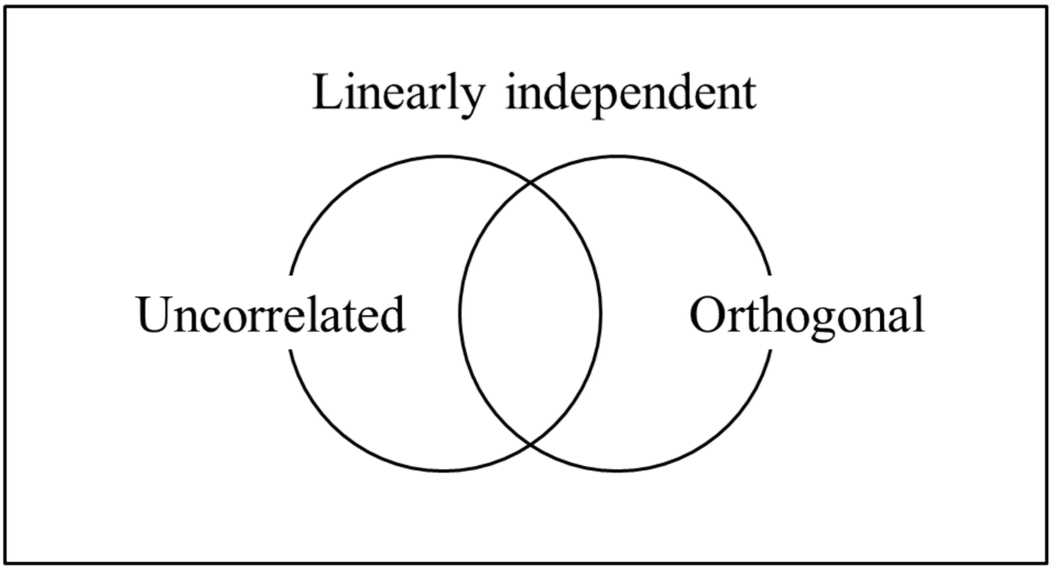

Figure 1. Venn diagram that represents the relationship between linearly independent, orthogonal and uncorrelated variables.

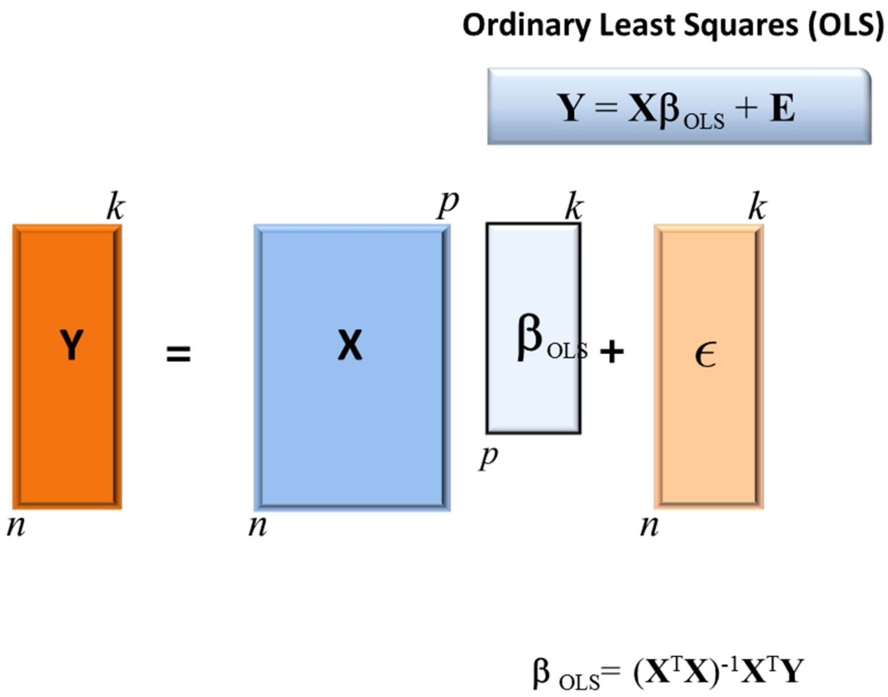

Figure 2, A. Geometrical shape of the ordinary least squares method. X, the matrix for the predictor variables; Y, the matrix for the response variables; E, the matrix of residuals for Y; β_{OLS}, the matrix solution of least squares and calculated by multivariate Gauss-Markov theorem.

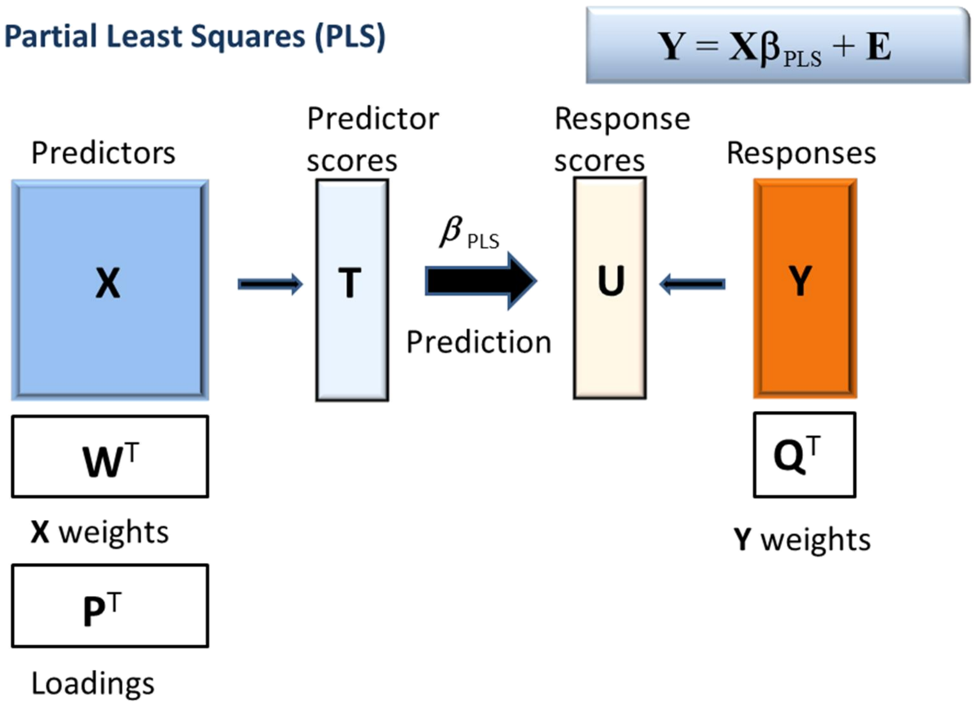

Figure 2, B. Geometrical shape of the partial least squares method. W^T, the matrix of weights for X; P^T, the matrix of loadings for X; T, the matrix of X scores; U, the matrix of Y scores; Q^T, the matrix of Y weights

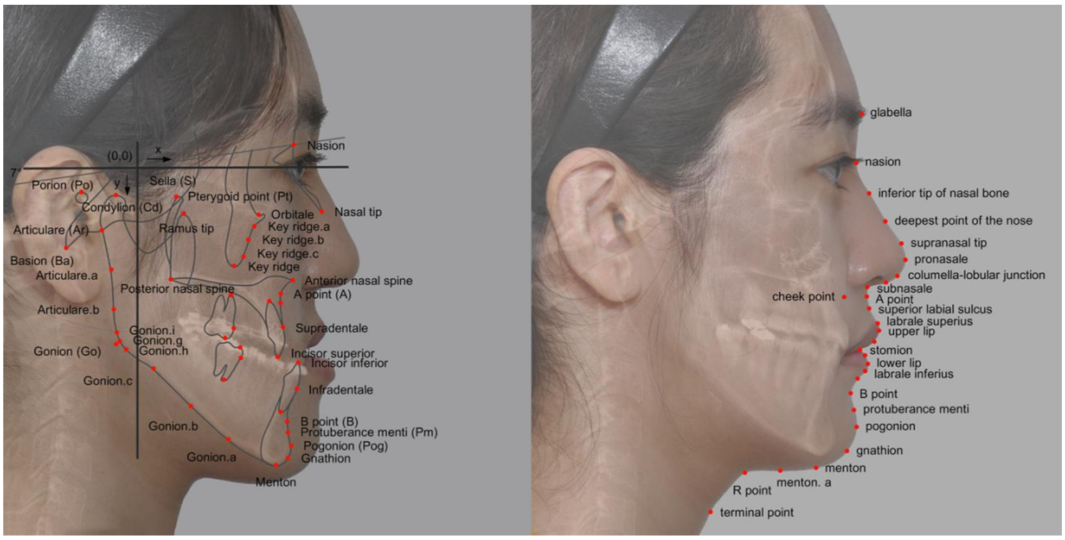

Figure 3. Diagram showing reference planes and cephalometric landmarks used in present study. ***Left,*** image composed from preoperative radiograph, with hard tissue landmarks given in capital letters. ***Right,*** soft tissue landmarks shown on the follow-up cephalogram and given in lowercase letters.

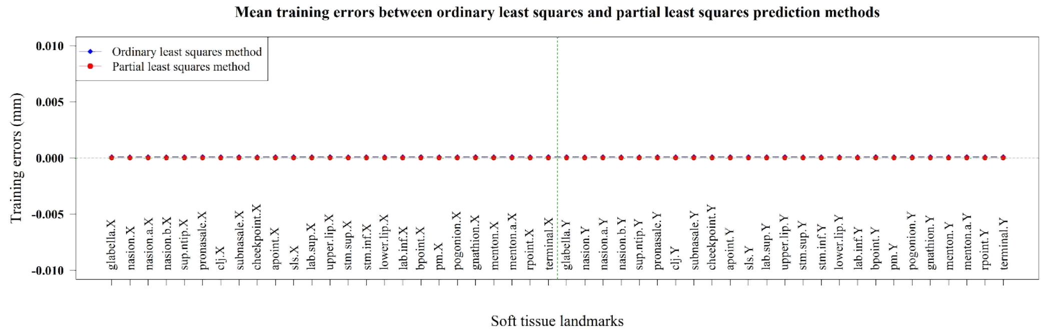

Figure 4, A. Mean training errors between ordinary least squares and pratial least squares prediction methods.

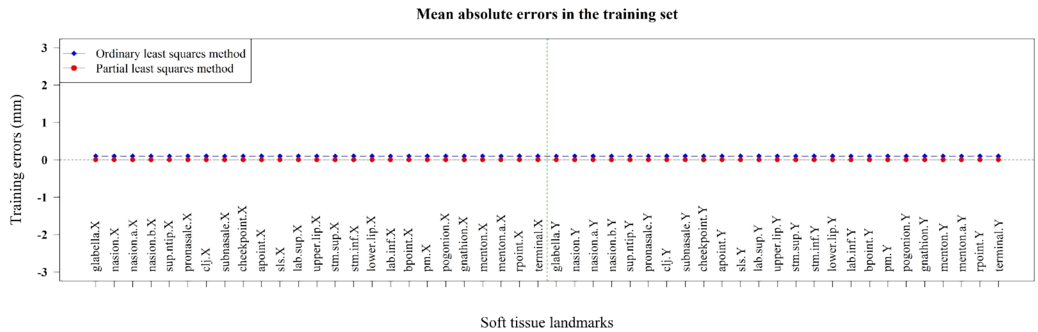

Figure 4, B. Mean absolute errors in the training set between ordinary least squares and partial least squares prediction.

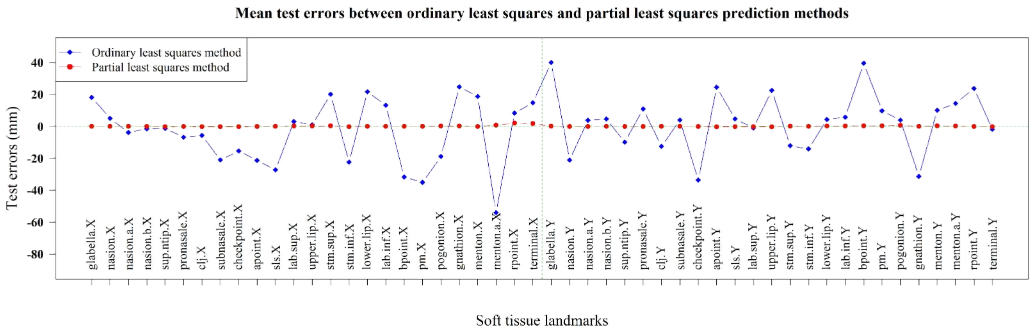

Figure 4, C. Mean test errors between ordinary least squares and partial least squares prediction.

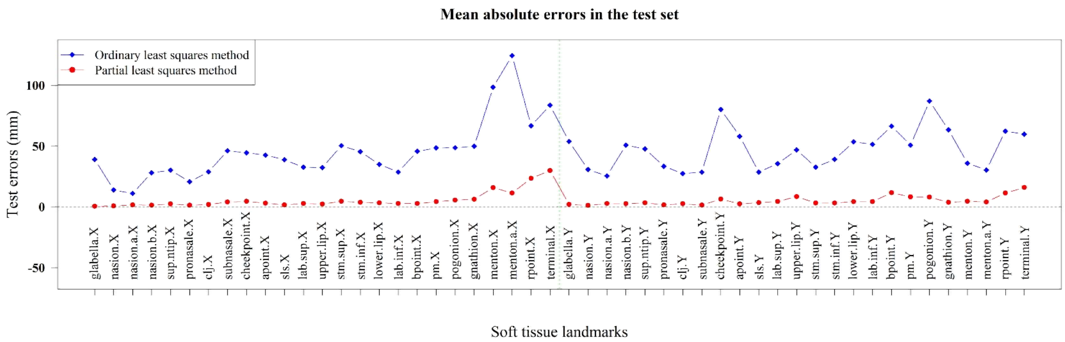

Figure 4, D. Mean absolute test errors in the test set between ordinary least squares and partial least squares prediction

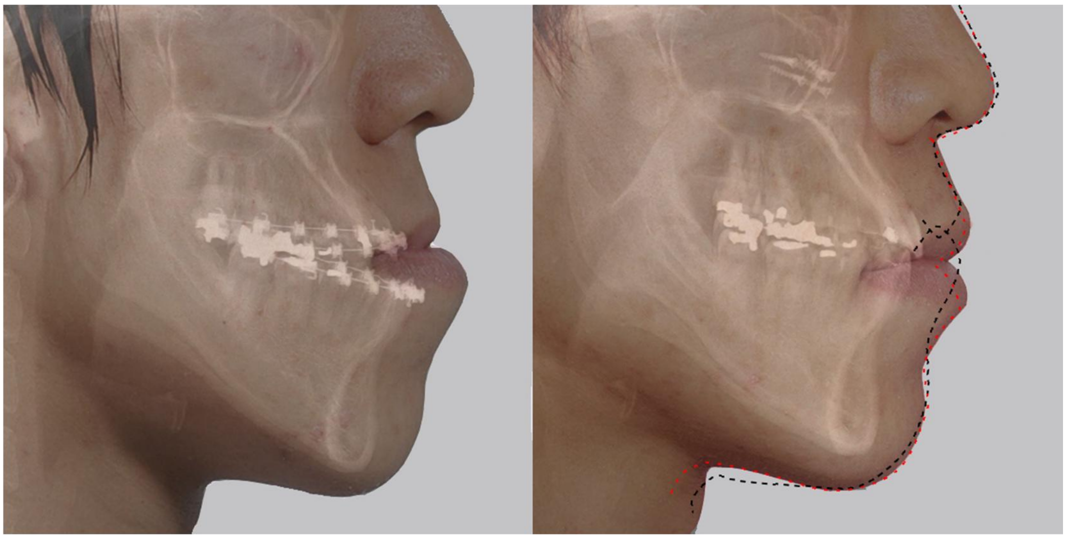

Figure 5, A. Patien id, 586812

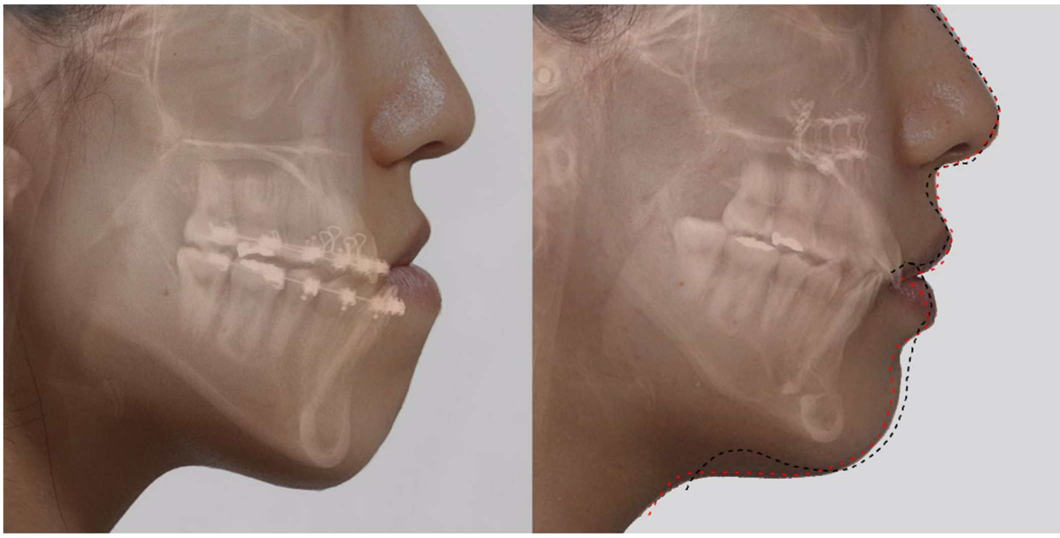

Figure 5, B. Patient id, 531771

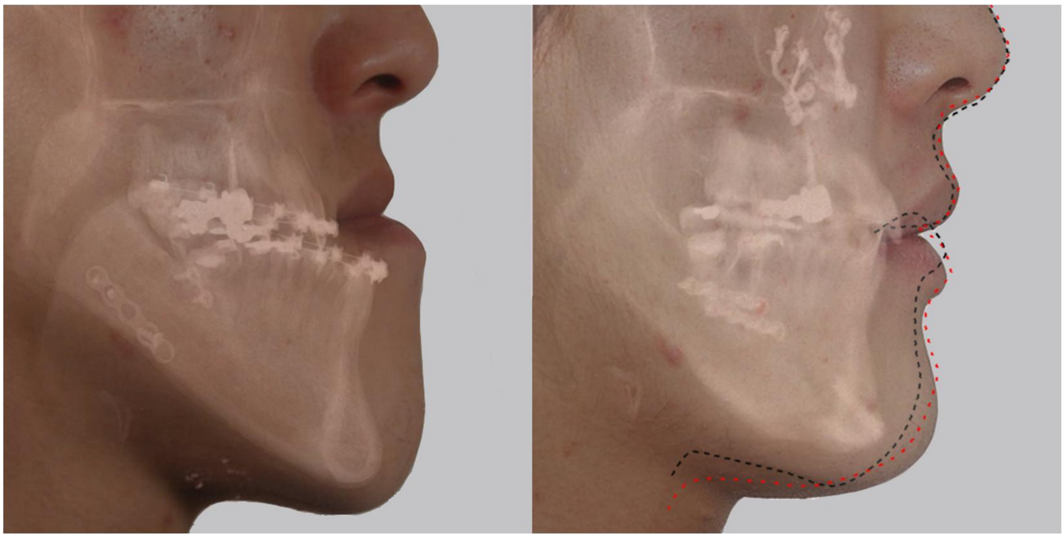

Figure 5, C. Patient id, 612444

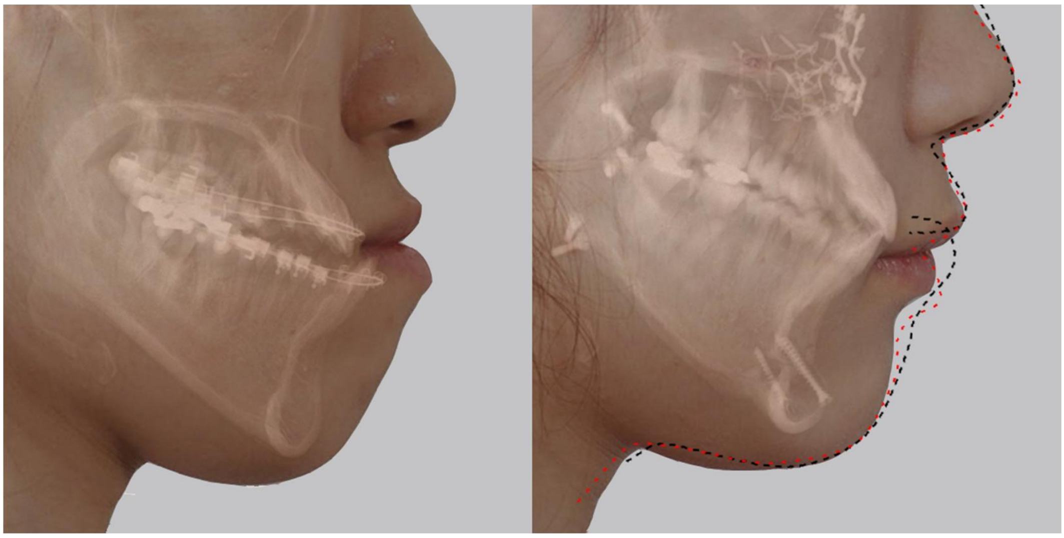

Figure 5, D. Patient id, 614721

Figure 5. Comparison the prediction accuracy between conventional method (using a commerical software; Vceph ver 4.3, Osstem, Korea) and partial least squares method in the clinical cases. *Left*, lateral photograph taken before surgery; *Right*, lateral photograph taken at the debonding stage. Black dash line, prediction by OLS method (VCeph); red dash line, prediction by PLS method.

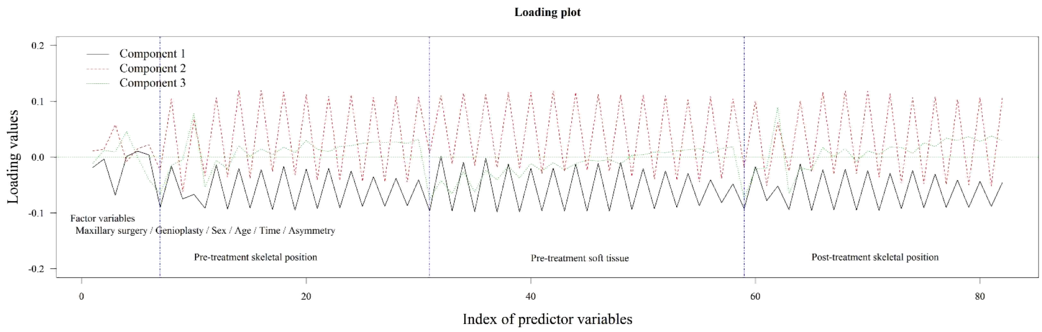

Figure 6. Loading plot of PLS method. The loading value indicates the magnitude of predictor variables in predicting the response. The number of components indicates the order of power of influence among PLS components. Component 1 is the most powerful component of the prediction model. Components are the latent variables that showed which of predictor variables played an important role in predicting the soft tissue position. Factor variables includ an existance of maxillary surgery, an existance of genioplasty, sex, age, time after surgery, absolute amount of asymmetry.

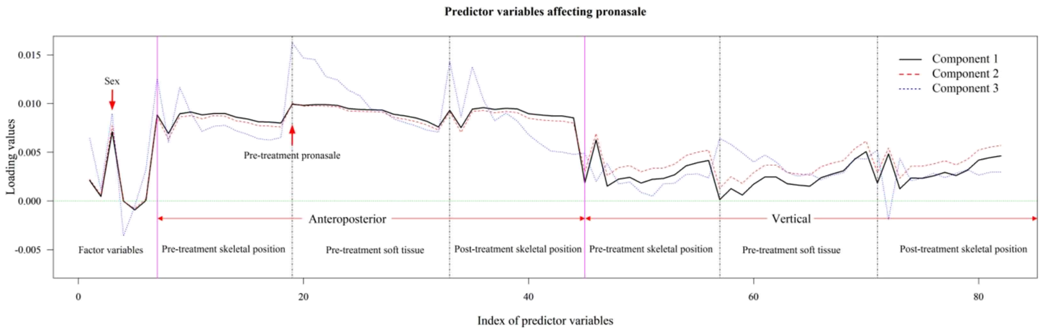

Figure 7. A loading plot of the fittted three major PLS components for anteroposterior position (x-axis) of the selected landmarks. Predictor variables were sorted by anteroposterior variables and vertical axes for convinience from the original loading plot **(Figure 6).**

A. Pronasale.

Among the factor variables, sex variable had an important role. Both the presurgical skeletal and soft tissue characteristics as well as the amount of surgical repositioning contributed to predicting the soft tissue response after surgery. It was obvious that when the response variable was an anteroposterior response, anteroposterior variables exerted more influence than the vertical variables. To restate, when predicting the nasal tip response in the x-axis, in addition to the anteroposterior variables, the vertical variables have participated to some extent. It was also notable that some neighboring soft tissue landmarks and all the skeletal landmarks as a whole had a greater influence on the predictions of specific soft tissue landmarks then the pre-surgical landmark of that actual soft tissue point itself (in this plot, pre-surgical pronasale).

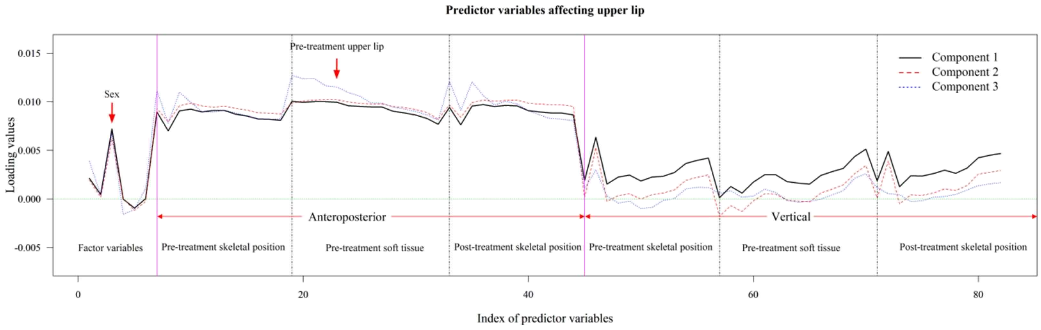

Figure 7, B. Upper lip.

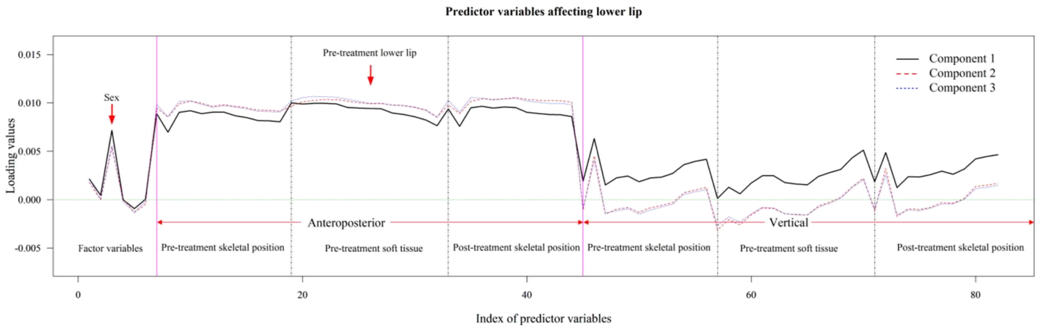

Figure 7, C. Lower lip.

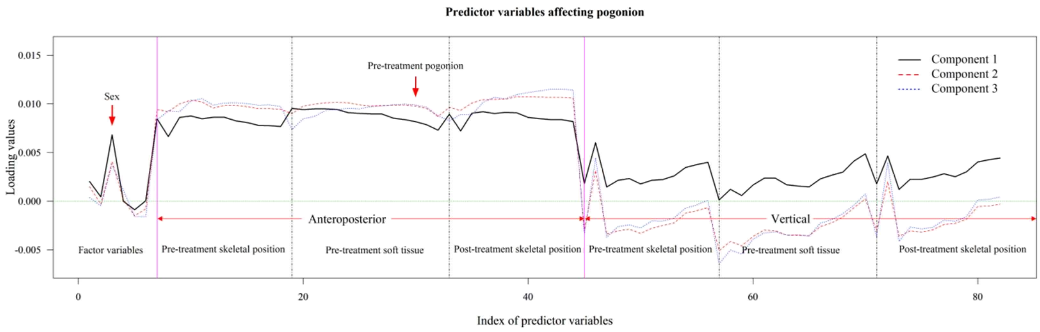

Figure 7, D. Pogonion.

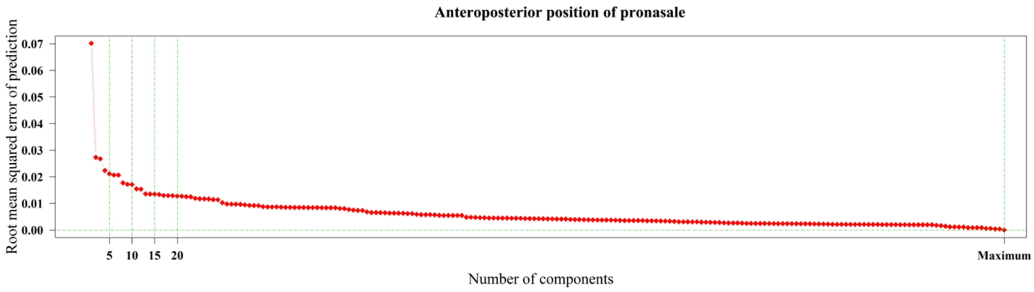

Figure 8, A. pronasale

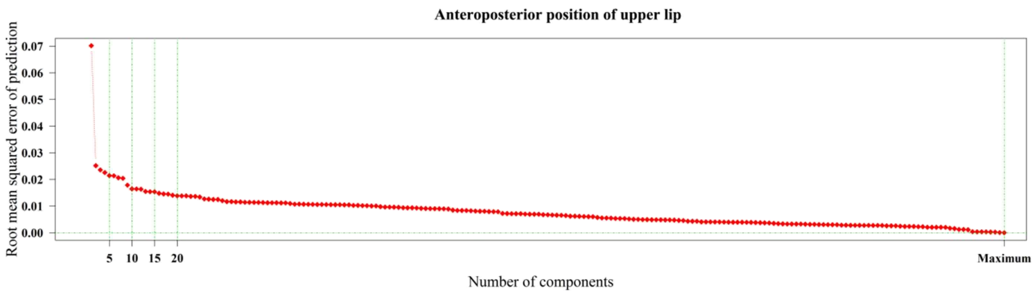

Figure 8, B. upper lip

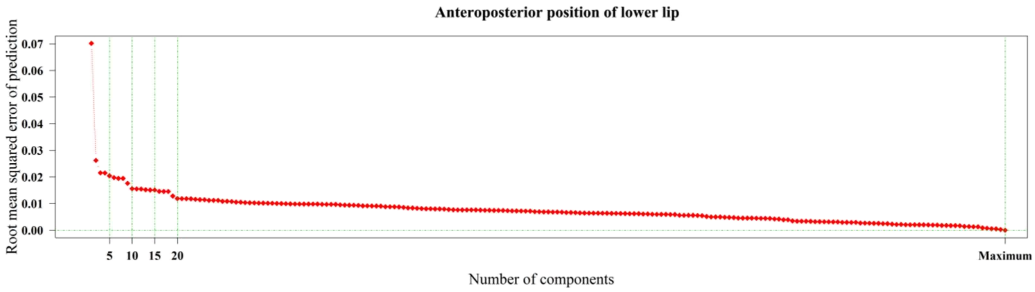

Figure 8, C. lower lip

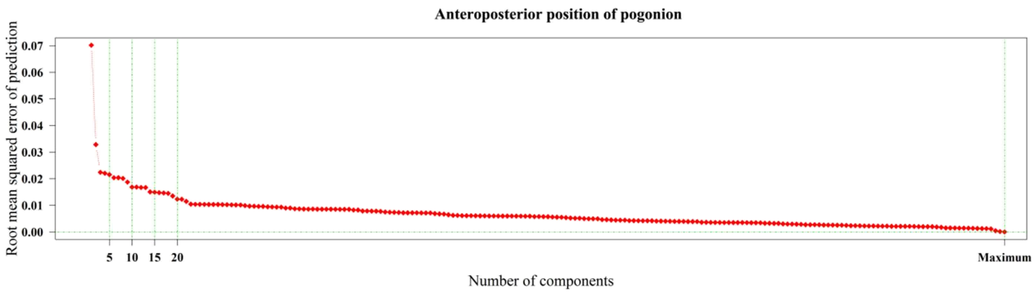

Figure 8, D. pogonion

Figure 8. Cross-validated RMSEP (root mean squared error of prediction) curve in the training set.

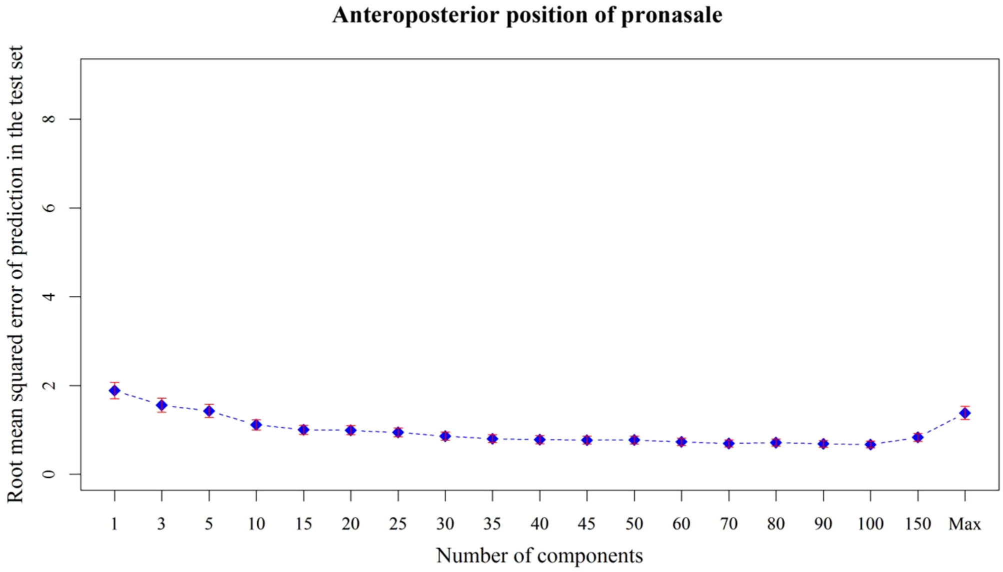

Figure 9, A. Anteroposterior position of pronasale

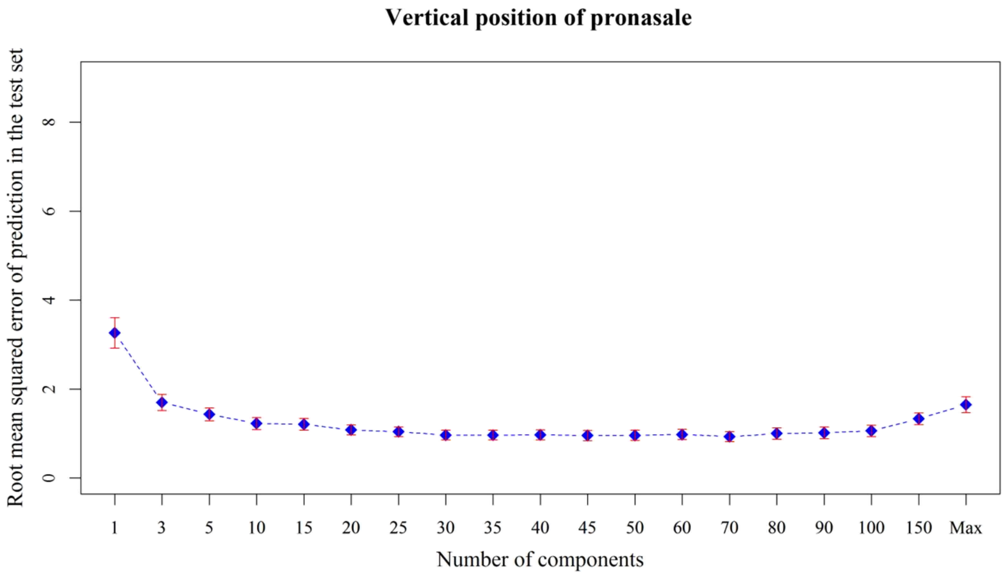

Figure 9, B. Vertical position of pronasale

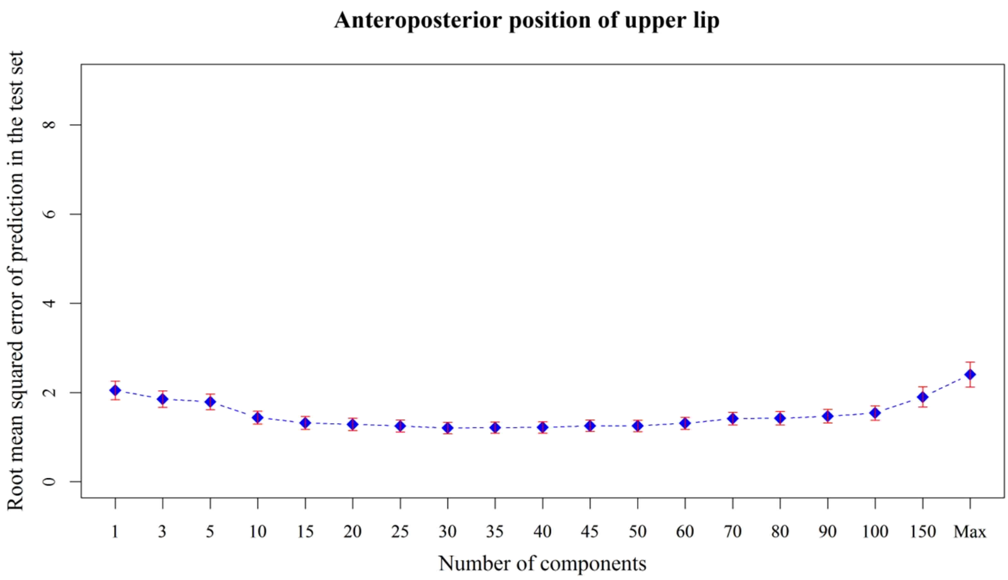

Figure 9, C. Anteroposterior position of upper lip

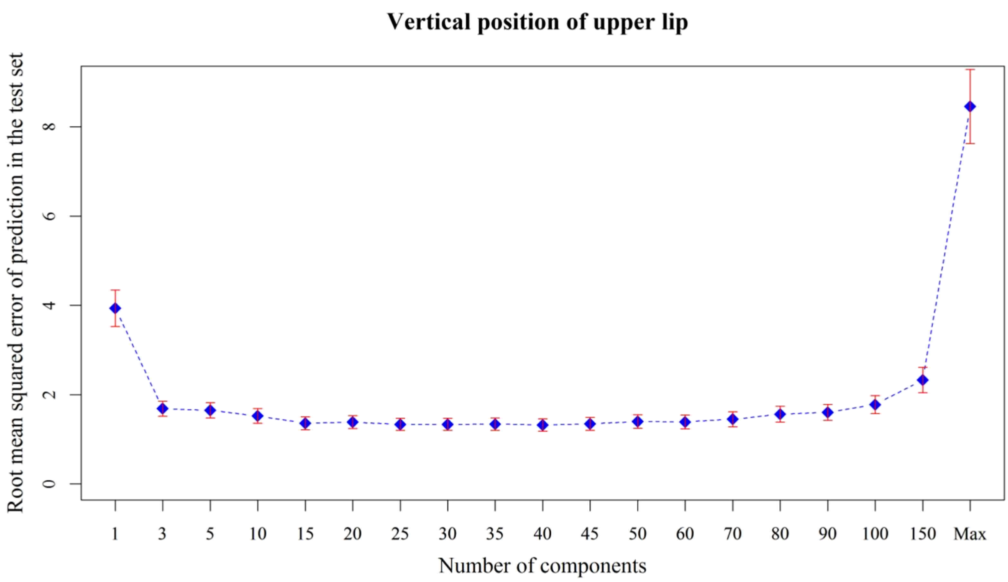

Figure 9, D. Vertical position of upper lip

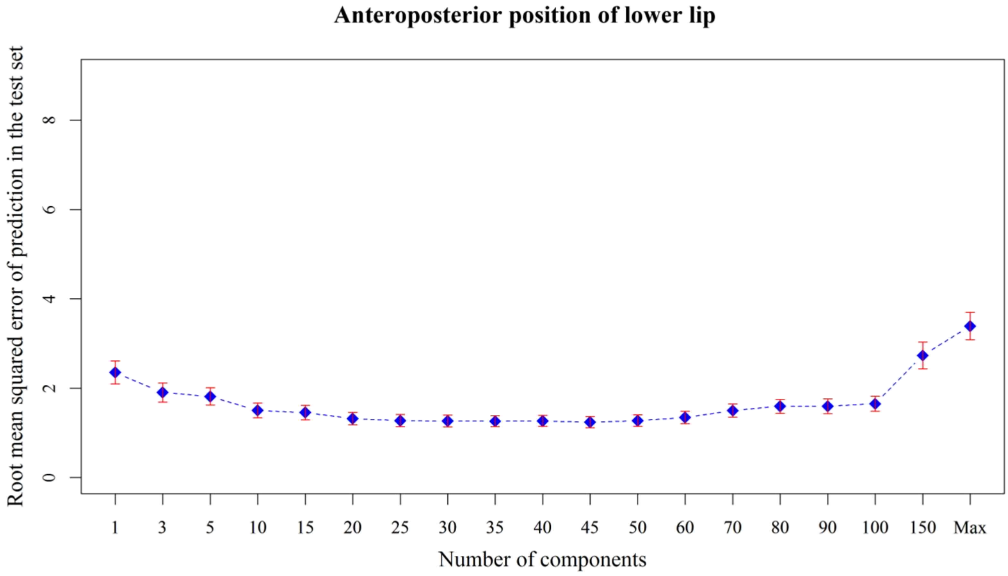

Figure 9, E. Anteroposterior position of lower lip

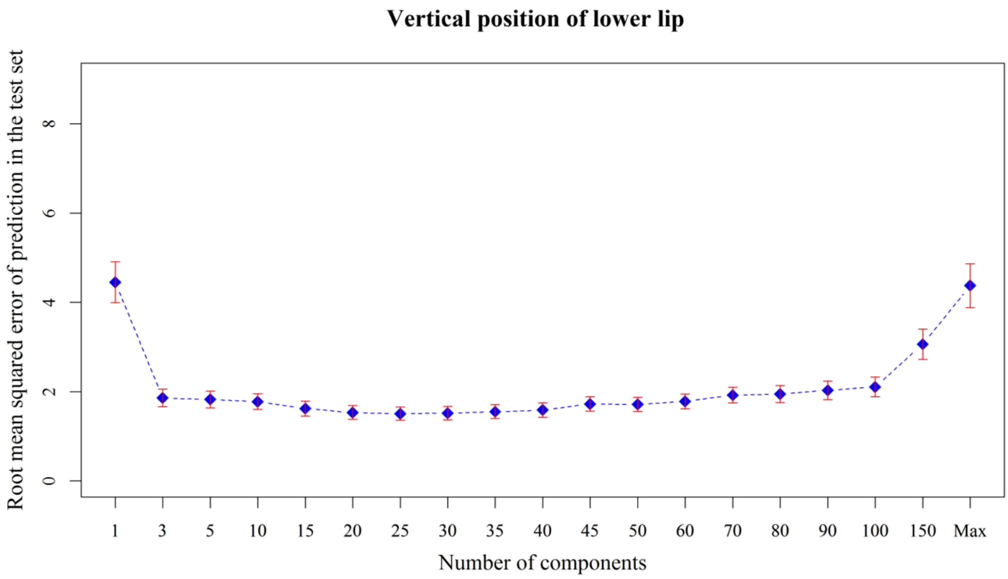

Figure 9, F. Vertical position of lower lip

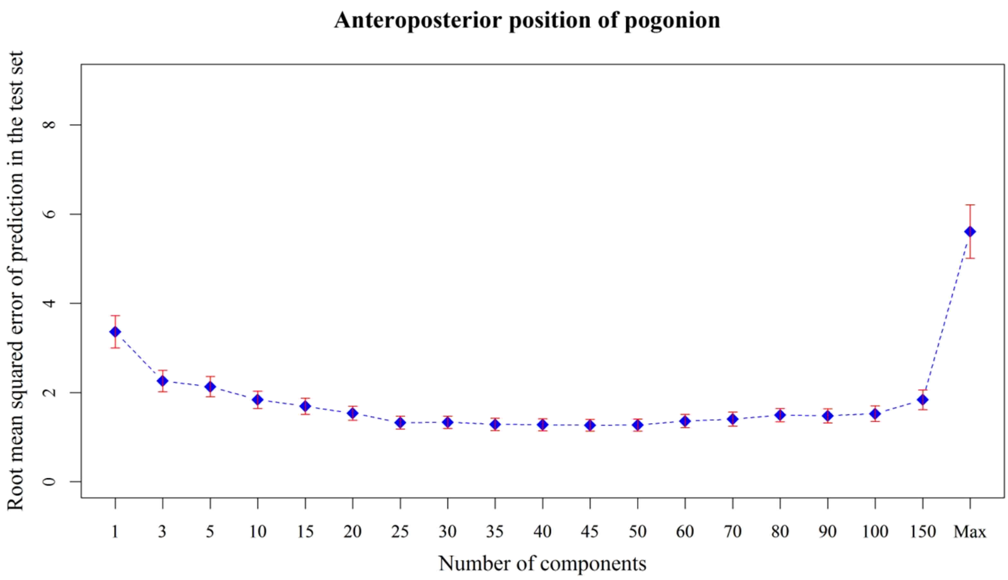

Figure 9, G. Anteroposterior position of pogonion

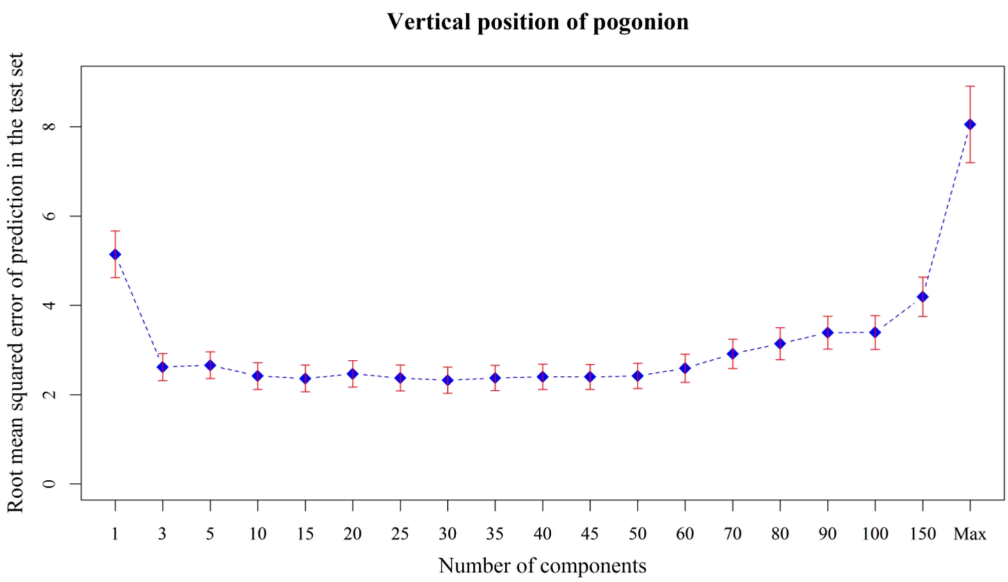

Figure 9, H. Vertical position of pogonion

Figure 9. Cross-validated RMSEP(root mean squared error of prediction) in the test set.

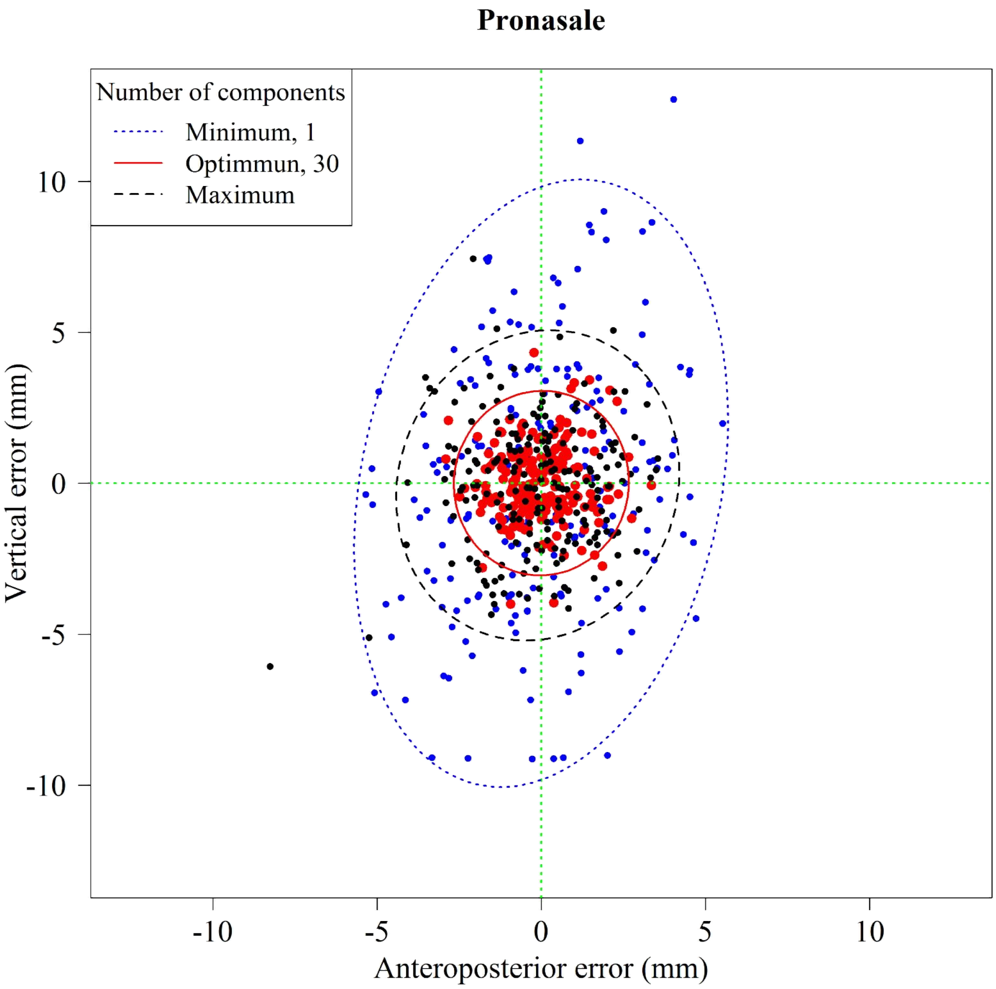

Figure 10, A. pronasale

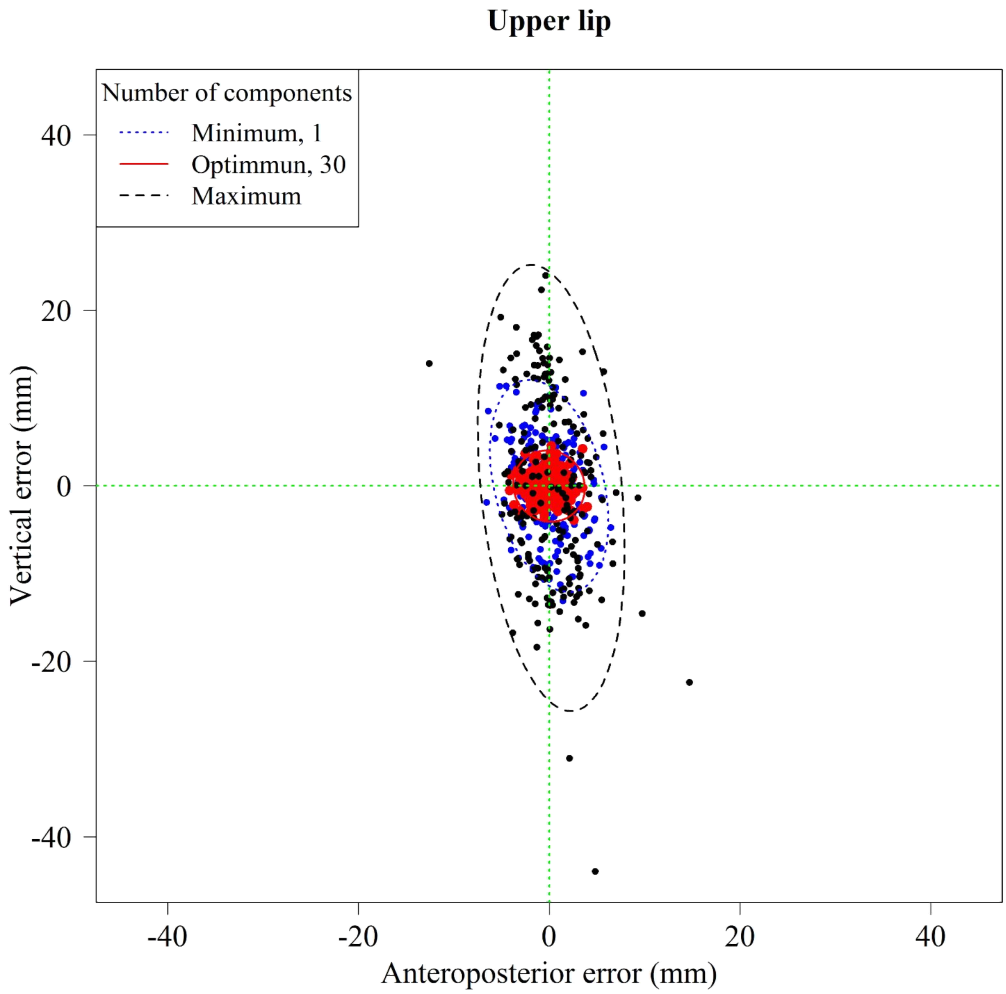

Figure 10, B Upper lip

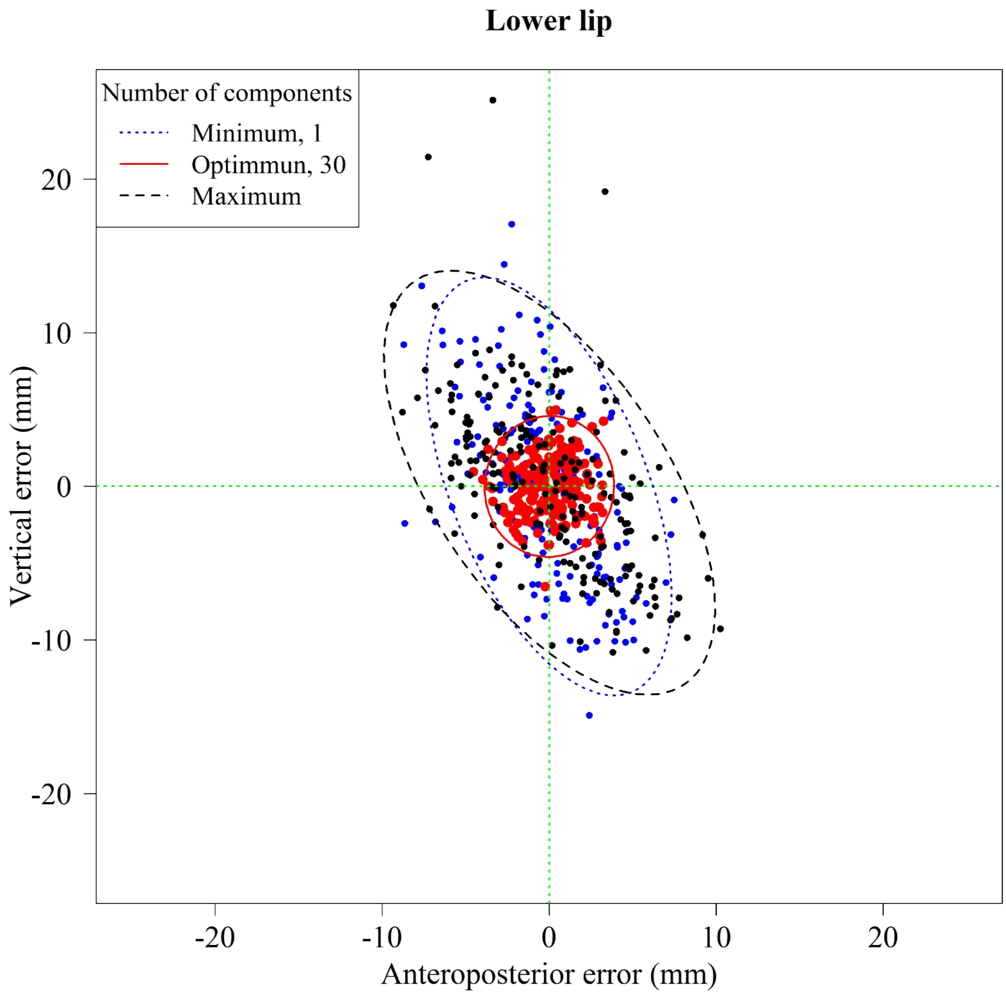

Figure 10, C. Lower lip.

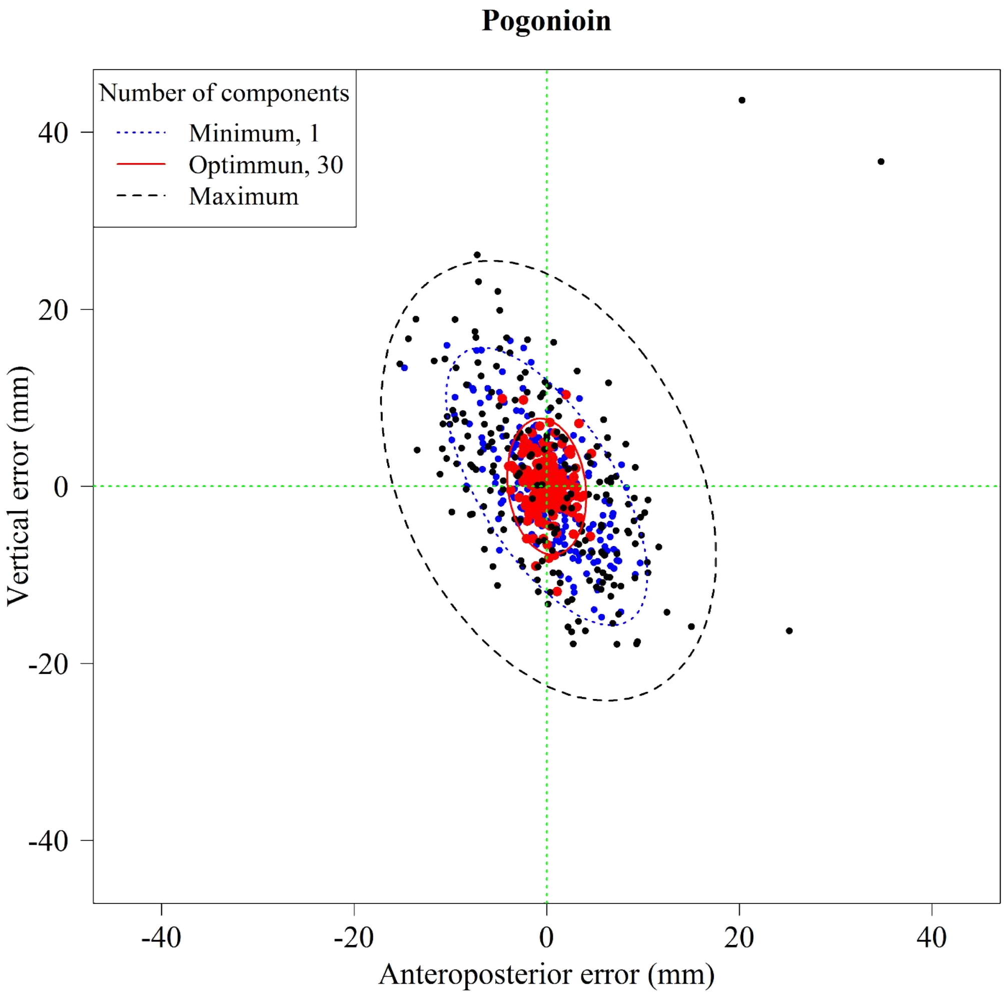

Figure 10, D. Pogonion

Figure 10. Scattergrams and 95% confidence ellipse for the bias that were obtained from the number of component was minimum (n=1), optimium (n=30), and maximum.

Table 1. The subjects′ sex, age, and other characteristics

Variables		N	Mean	SD	Min	Max
Age (years)						
	Female	103	23.8	5.1	16.0	50.5
	Male	101	23.6	3.5	18.8	39.1
Time after surgery (months)						
			9.1	3.9	3.7	29.4
Type of maxillary surgery						
	No	71				
	Yes	133				
Additional genioplasty						
	No	123				
	Yes	81				
Asymmetry (mm)						
	None	63				
	Asymmetry (absolute value)	141	4	1.8	1	10
Overjet before surgery (mm)		204	-5.8	3.8	-19.6	2.2
Overbite before surgery (mm)		204	-0.2	1.8	-5.4	5.9
Amount of surgical repositioning at point A (mm)		133				
	Anteroposterior repositioning		1.4	1.9	-4.4	7
	Vertical repositioning		-1.1	2.3	-7.1	4.4
Amount of surgical repositioning at point B (mm)		204				
	Anteroposterior repositioning		-7.3	3.8	-24.1	4.1
	Vertical repositioning		-2.9	4.2	-14.8	11.4

Table 2-1. Comparison of the soft tissue prediction errors between conventional ordinary least square (OLS) and partial least squares (PLS) prediction methods (*x*-axis)

Variable or coordinate	Bias		*P* -value	Mean absolute error		*P* -value
	OLS	PLS		OLS	PLS	
Horizontal (*x* position [mm])						
glabella	18.12	-0.03	0.4867	39.03	0.54	0.1393
nasion	5.01	0.00	0.2565	13.92	0.77	0.0026†
inferior tip of nasal bone	-3.82	-0.01	0.1936	11.05	1.73	0.0012†
deepest point of the nose	-1.51	-0.05	0.8969	28.10	1.38	0.0171*
supranasal tip	-1.29	-0.17	0.9051	30.20	2.53	0.0029†
pronasale	-6.80	-0.11	0.4050	20.73	1.38	0.0151*
columella-lobular junction	-5.66	-0.19	0.5477	28.83	2.00	0.0028†
subnasale	-21.04	-0.14	0.2699	46.21	4.04	0.0248*
cheek point	-15.39	-0.24	0.3638	44.52	4.54	0.0152*
soft tissue A point	-21.32	-0.07	0.2468	42.61	3.10	0.0305*
superior labial sulcus	-27.21	-0.02	0.0804	38.81	1.65	0.0162*
labrale superius	2.97	0.21	0.7741	32.66	2.75	0.0015†
upper lip	0.98	0.21	0.9346	32.32	2.40	0.0010†
stomion	20.07	0.27	0.3819	50.38	4.53	0.0414*
lower lip	21.69	0.02	0.1146	34.96	3.39	0.0209*
labrale inferius	13.19	0.05	0.1154	28.67	2.77	0.0016†
soft tissue B point	-31.69	0.02	0.2009	45.73	2.72	0.0819
protuberance menti	-35.05	0.06	0.1634	48.51	4.36	0.0789
pogonion	-18.87	0.20	0.3977	48.69	5.61	0.0550
gnathion	24.74	0.20	0.2790	49.92	6.26	0.0524
menton	18.73	-0.07	0.6302	98.46	15.80	0.0328*
menton. a	-54.00	0.77	0.3285	124.45	11.47	0.0429*
R point	8.33	2.07	0.7201	66.75	23.50	0.0105*
terminal point	14.82	1.75	0.6262	83.63	29.96	0.0417*

* $P < 0.05$, † $P < 0.01$, ‡ $P < 0.001$, result of paired *t* test.

Table 2-2. Comparison of the soft tissue prediction errors between conventional OLS and PLS prediction methods (y-axis)

Variable or coordinate	Bias		P-value	Mean absolute error		P -value
	OLS	PLS		OLS	PLS	
Vertical (y position [mm])						
glabella	39.94	0.07	0.2754	53.8	2.05	0.1562
nasion	-21.09	-0.1	0.1966	30.72	1.25	0.0692
inferior tip of nasal bone	3.8	-0.04	0.5609	25.5	2.73	<0.0001‡
deepest point of the nose	4.65	0.06	0.8118	50.79	2.6	0.0116*
supranasal tip	-9.87	-0.04	0.5038	47.72	3.34	0.0022†
pronasale	10.91	-0.07	0.4761	33.36	1.65	0.0386*
columella-lobular junction	-12.54	0.02	0.1447	27.38	2.64	0.0037†
subnasale	3.95	0.04	0.7376	28.59	1.49	0.0194*
cheek point	-33.62	-0.04	0.2219	80.13	6.37	0.0067†
soft tissue A point	24.5	-0.26	0.3048	58.02	2.48	0.0203*
superior labial sulcus	4.69	-0.23	0.5298	28.66	3.5	0.0011†
labrale superius	-0.99	-0.22	0.9366	35.58	4.49	<0.0001‡
upper lip	22.52	-0.25	0.1091	46.83	8.45	0.0060†
stomion	-12.1	0.11	0.2565	32.66	3.26	0.0056†
lower lip	4.31	0.24	0.8403	53.48	4.38	0.0142*
labrale inferius	5.79	0.24	0.7247	51.47	4.24	0.0024†
soft tissue B point	39.53	0.36	0.1337	66.37	11.72	0.0347
protuberance menti	9.73	0.3	0.5807	50.81	8.2	0.0115*
pogonion	3.89	0.64	0.9321	87.09	8.05	0.0366*
gnathion	-31.35	0.06	0.3564	63.43	3.79	0.0787
menton	10.09	0.28	0.1801	35.91	4.65	<0.0001‡
menton. a	14.39	0.17	0.0683	30.33	4.08	<0.0001‡
R point	23.73	-0.04	0.176	62.33	11.41	0.0033†
terminal point	-1.86	-0.18	0.8783	59.77	15.95	<0.0001‡

* $P < 0.05$, † $P < 0.01$, ‡ $P < 0.001$, result of paired t test.

Table 3. Comparisons of the prediction errors depending on the number of components

Number of components	Root mean squared error of prediction in the test set																	
	1	3	5	10	15	20	25	30	35	40	45	50	60	70	80	90	100	150	Max
Horizontal (x position [mm])																			
pronasale	1.89	1.56	1.43	1.11	1.00	0.99	0.94	0.86	0.80	0.78	0.77	0.77	0.73	0.70	0.71	0.68	0.67	0.83	1.38
upper lip	2.05	1.85	1.79	1.44	1.32	1.29	1.25	1.21	1.21	1.22	1.26	1.25	1.31	1.42	1.42	1.47	1.54	1.90	2.40
lower lip	2.35	1.90	1.82	1.50	1.46	1.32	1.28	1.27	1.26	1.27	1.24	1.28	1.35	1.50	1.60	1.60	1.65	2.73	3.39
pogonion	3.36	2.26	2.13	1.84	1.69	1.54	1.33	1.34	1.29	1.28	1.27	1.27	1.36	1.41	1.50	1.48	1.53	1.84	5.61
Vertical (y position [mm])																			
pronasale	3.27	1.70	1.44	1.23	1.21	1.08	1.04	0.97	0.97	0.97	0.96	0.96	0.98	0.93	1.00	1.02	1.06	1.33	1.65
upper lip	3.94	1.69	1.65	1.53	1.36	1.39	1.34	1.33	1.34	1.32	1.35	1.40	1.39	1.45	1.56	1.61	1.78	2.33	8.45
lower lip	4.45	1.86	1.82	1.78	1.62	1.53	1.51	1.52	1.55	1.59	1.72	1.72	1.78	1.92	1.95	2.03	2.11	3.06	4.38
pogonion	5.14	2.62	2.66	2.42	2.36	2.47	2.38	2.32	2.37	2.40	2.40	2.42	2.59	2.91	3.14	3.39	3.39	4.19	8.05

Table 4-1. Comparison of the test errors between 1-jaw (mandibular setback) and 2-jaw surgery patients

Variable or coordinate	Bias		*P*-value	Mean absolute error		*P*-value
	1-jaw	2-jaw		1-jaw	2-jaw	
Horizontal (*x* position [mm])						
glabella	0.001	-0.034	0.859	1.126	0.969	0.173
nasion	-0.053	0.024	0.589	0.703	0.766	0.503
inferior tip of nasal bone	-0.114	0.049	0.316	0.89	0.791	0.333
deepest point of the nose	-0.098	0.043	0.377	0.764	0.880	0.277
supranasal tip	-0.123	0.034	0.397	0.879	1.158	0.017*
pronasale	-0.032	0.023	0.712	0.738	0.927	0.037*
columella-lobular junction	-0.033	0.023	0.777	1.045	1.027	0.881
subnasale	-0.014	-0.004	0.959	0.998	0.911	0.440
cheek point	-0.216	0.068	0.264	1.432	1.221	0.176
soft tissue A point	-0.067	0.017	0.612	0.858	0.923	0.517
superior labial sulcus	-0.064	0.012	0.649	0.847	1.034	0.059
labrale superius	-0.058	0.005	0.751	1.017	1.228	0.072
upper lip	-0.063	0.008	0.736	1.066	1.281	0.092
stomion	0.029	-0.026	0.837	1.43	1.538	0.491
lower lip	0.105	-0.049	0.513	1.286	1.255	0.828
labrale inferius	0.082	-0.044	0.573	1.205	1.181	0.857
soft tissue B point	0.083	-0.047	0.544	1.135	1.255	0.346
protuberance menti	0.104	-0.059	0.465	1.17	1.232	0.649
pogonion	0.094	-0.054	0.553	1.359	1.325	0.822
gnathion	0.034	-0.007	0.904	1.916	1.742	0.401
menton	0.208	-0.092	0.702	3.961	4.164	0.692
menton.a	0.285	-0.046	0.645	4.083	3.346	0.098
R point	0.487	-0.080	0.554	5.058	4.320	0.252
terminal point	0.593	-0.171	0.507	6.350	5.240	0.135

* $P < 0.05$, result of *t* test.

Table 4-2. Comparison of the test errors between 1-jaw and 2-jaw surgery

Variable or coordinate	Bias		*P*-value	Mean absolute error		*P*-value
	1-jaw	2-jaw		1-jaw	2-jaw	
Vertical (*y* position [mm])						
glabella	-0.069	0.036	0.590	0.931	0.734	0.178
nasion	-0.011	-0.010	0.993	0.748	0.592	0.088
inferior tip of nasal bone	-0.084	0.023	0.517	0.790	0.879	0.422
deepest point of the nose	-0.027	-0.005	0.923	1.001	1.250	0.107
supranasal tip	-0.089	-0.002	0.759	1.298	1.468	0.401
pronasale	0.040	-0.014	0.776	1.074	0.910	0.159
columella-lobular junction	-0.003	0.015	0.773	0.843	0.880	0.686
subnasale	-0.049	0.029	0.576	0.698	0.856	0.057
cheek point	0.005	-0.017	0.953	1.840	2.080	0.315
soft tissue A point	-0.002	0.034	0.863	1.052	1.188	0.265
superior labial sulcus	0.002	-0.011	0.963	1.499	1.400	0.535
labrale superius	-0.096	0.004	0.697	1.380	1.346	0.828
upper lip	-0.075	-0.001	0.762	1.370	1.315	0.698
stomion	-0.054	0.019	0.687	0.960	1.000	0.717
lower lip	-0.063	0.018	0.777	1.598	1.476	0.471
labrale inferius	-0.039	0.023	0.850	1.849	1.581	0.196
soft tissue B point	-0.054	0.033	0.808	1.932	1.921	0.962
protuberance menti	-0.034	-0.016	0.961	1.787	2.051	0.258
pogonion	-0.052	-0.028	0.958	2.083	2.454	0.208
gnathion	-0.039	-0.029	0.974	1.685	1.581	0.625
menton	0.089	-0.058	0.616	1.609	1.453	0.390
menton.a	0.171	-0.061	0.569	2.122	2.172	0.844
R point	0.023	0.030	0.992	3.576	3.756	0.690
terminal point	0.475	-0.129	0.480	4.166	4.135	0.959

* $P < 0.05$, result of t test.